MASTER

Study of a Delayed McKean-Vlasov process

van Schooten, Bart

Award date:
2021

Study of a Delayed McKean-Vlasov process

Master Thesis
Industrial and Applied Mathematics
Technische Universiteit Eindhoven

Bart van Schooten

Supervisors:
Alberto Chiarini
Markus Fischer

17th August 2021

Contents

Glossary

Throughout this thesis, we use the following acronyms:

a.s.	almost surely
i.i.d.	independent and identically distributed
LDP	large deviation principle
ODE	ordinary differential equation
SDE	stochastic differential equation

For mathematical notations, we use the following definitions:

$\varnothing$	the empty set
$\mathbb{1}$	the indicator function
$a \wedge b$	the minimum of a and b
$a \vee b$	the maximum of a and b
$\mathbb{R}$	the real line
$\mathbb{R}_+$	the set of non-negative numbers, i.e. $\mathbb{R}_+ = [0, \infty)$
$\mathbb{R}^d$	the d-dimensional Euclidean space
$\mathbb{R}^{d \times m}$	the space of real $d \times m$-matrices
$\lvert x \rvert$	the Euclidean norm of a vector x
$B(x_0, R)$	the open ball with radius R around x_0, i.e. $B(x_0, R) = \{x : \lvert x - x_0 \rvert < R\}$
$\langle x, y \rangle$	the inner product (scalar product) of the vectors x and y
$\mathbb{E}[X]$	the expected value of X
$\mathbb{P}(A)$	the probability of the event A
$C(\mathbf{X}, \mathbf{Y})$	the family of continuous functions from $\mathbf{X}$ to $\mathbf{Y}$
$L^p([a,b], \mathbb{R}^d)$	the family of functions $h : [a,b] \to \mathbb{R}^d$ such that $\int_a^b \lvert h(t) \rvert \,\mathrm{d}t < \infty$
$\mathcal{L}^p([a,b], \mathbb{R}^d)$	the family of $\mathbb{R}^d$-valued random processes $\{f(t)\}_{a \le t \le b}$ such that $\int_a^b \lvert f(t) \rvert^p \,\mathrm{d}t < \infty$ a.s.
$\mathcal{M}^p([a,b], \mathbb{R}^d)$	the family of $\mathbb{R}^d$-valued random processes $\{f(t)\}_{a \le t \le b}$ such that $\mathbb{E}\left[\int_a^b \lvert f(t) \rvert^p \,\mathrm{d}t\right] < \infty$

Instead of writing $\{f(t)\}_{a \le t \le b}$ we may also write $\{f(t)\}$ or simply f when it is clear from the context on which intervals the process is defined. Definitions not explained here will be explained when they first appear.

1. Introduction

Imagine a boat with people on board moving around and a captain trying to keep the boat balanced. Suppose that the movement of the people can be described by their current position on the boat with the addition of some noise, as their behaviour can not be fully determined. The captain, trying to keep the boat in balance, will tell all people, regardless of their own position, to move in a certain direction based on which direction the boat is currently leaning. He, however, only notices that the boat is leaning in a certain direction with some delay. If we wish to describe this model mathematically, we could use *stochastic delay equations*. Let us explain how to model this scenario. For simplicity we will assume that the boat stretches out indefinitely in all directions, such that we do not have to take into account boundary conditions. Let N denote the amount of people on the boat. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be a function describing in which direction a person moves based on their current position, and let $g : \mathbb{R}^2 \to \mathbb{R}^2$ be a function describing how a person moves based on the instructions of the captain, which bases his instructions on the center of mass of all the people on the boat with a delay $\delta > 0$. Let $X^i(t)$ denote the position of the i'th person at time t. The following *stochastic differential equation (SDE)* now describes the movement of the people on the ship:

$$\mathrm{d}X^i(t) = f(X^i(t))\mathrm{d}t + g\left(\frac{1}{N}\sum_{j=1}^{N} X^j(t-\delta)\right)\mathrm{d}t + \varepsilon \mathrm{d}B^i(t), \quad i = 1, \dots, N. \tag{1.1}$$

Here $B^i, i = 1, \dots, N$ denote independent Brownian motions, modeling the uncertainty of the movement of the people, and $\varepsilon > 0$ is a parameter scaling this uncertainty. In order to complete this model we still need to give the appropriate initial conditions, but we will do this in Section 1.1. An interesting question to ask is what happens when the amount of people on the boat becomes large. The law of large numbers suggest that the empirical mean converges towards the expected value of the random variables. We therefore also consider the *delayed McKean-Vlasov equation*

$$\mathrm{d}X(t) = f(X(t))\mathrm{d}t + g\left(\mathbb{E}\left[X(t-\delta)\right]\right)\mathrm{d}t + \varepsilon \mathrm{d}B(t). \tag{1.2}$$

These special types of SDEs, where the evolution of the process depends not only on the state of the process but also on its law, were first studied by McKean (1966).

An interesting question is whether the ship will eventually stabilise, or if the delay will cause the ship to keep shifting from one side to the other over and over again. We briefly study this question for the deterministic equation ($\varepsilon = 0$) in Section 6.2.2. We find that the limiting behaviour highly depends on the choices for f and g and parameters of the model. During our research for this thesis we also considered many different other scenarios for both the deterministic equation and the SDEs. Our results however were not clear, and we were unsure what effects should be attributed to the model and which were nothing but artifacts of our approximation method. We therefore decided to not include these result in this thesis, and leave this question open for further research.

The model can also be given a more general interpretation. Consider N agents, whose current state can be described as a vector in $\mathbb{R}^d, d \geq 1$. Suppose that the evolution of these agents can be described as a function of their current *position* combined with a noise term. Suppose further that there exists a controlling agent which influences the evolution of all the agents based on their center of mass with a delay δ. We again arrive at Equation 1.1.

Let us give an example. Consider a finite-horizon problem. Let $X^i \in \mathbb{R}$ denote the amount of money a company receives from the government at time $T > 0$, where a positive number indicates that the company will receive money, while a negative number indicates that the company has to pay money to the government. In this case the controlling agent is the government itself, which tries to keep its expenses balanced, i.e. they strive to keep the mean payout 0. The function g in this case could for example be $g(x) = -x$, which pushes the mean towards 0. The function f would describe the intentions of the companies, which do not wish to pay money.

Besides its applications, the model is also interesting to study from a mathematical point of view. While the study of both stochastic delay equations and McKean-Vlasov equations is rich, the study of the combinations of the two seems much slimmer. In particular, we could not find any work that studies these kind of equations under the same assumptions as we will be making in this thesis.

1.1. Main model

We now describe the model in more detail. Let $d \geq 1$ denote the dimension of both the Brownian motion and the processes themselves. In a more general case it is also possible to let the Brownian motion be of a different dimension than the processes, but we will not consider that case in this thesis. Let $N \geq 1$ denote the amount of particles. We now have that $f : \mathbb{R}^d \to \mathbb{R}^d$ and $g : \mathbb{R}^d \to \mathbb{R}^d$ to match the dimensions of the process. Let $\delta > 0$ denote the delay, let $\varepsilon > 0$ denote the intensity of the noise and let $T > 0$ denote the time horizon on which we consider the problem. We consider the *interacting particle process* governed by Equation (1.1) on the interval $[0, T]$ with initial conditions

$$\begin{aligned} X^i(0) &= x_0^i, \\ X^i(s) &= \xi_s^i, \quad s \in [-\delta, 0), \quad i = 1, \dots, N. \end{aligned}$$

In general x_0^i and $\xi_s^i, s \in [-\delta, 0)$ are random variables taking values in $\mathbb{R}^d$. We also consider the McKean-Vlasov process, which consists of the Equation (1.2) on the interval $[0, T]$ and initial condition

$$\begin{aligned} X(0) &= x_0, \\ \mathbb{E}\left[X(s)\right] &= \xi_s, \quad s \in [-\delta, 0). \end{aligned}$$

Here we again have that x_0 is a random variable taking values in $\mathbb{R}^d$. This time, however, we have that $\xi : [-\delta, 0) \to \mathbb{R}^d$ is a deterministic function. This is because the interacting with the past only happens through the mean, so it is sufficient to only specify the mean of the process on this interval.

1.2. Main results

We just presented the McKean-Vlasov process as the limiting process of the interacting particle process when the amount of particles N is send to infinity. While intuitively this seems like a natural limit, it is not directly clear in what sense this limit should hold, if it does. In Chapter 4 we investigate this relation further. We find that under certain assumptions the McKean-Vlasov process can indeed be seen as the limiting process of the interacting particle process. If we couple the Brownian motions and the initial conditions, i.e. $B^1 = B$, $x_0^1 = x_0$ and $\mathbb{E}\left[\xi_s^1\right] = \xi_s, s \in [-\delta, 0)$, we find that the expected distance between the trajectories of X^1 and X in the supremum norm converges to 0 when N goes to infinity. We also obtain an upper-bound on the rate of convergence. In particular, we show, in Theorem 4.3, a result of the form

$$\sup_N N\mathbb{E}\left[\sup_{0 \leq t \leq T} |X^{1,N} - X|^2\right] < \infty.$$

In fact, the above result can be generalised even further. Assume that the initial conditions are independent. By considering n independent copies of the McKean-Vlasov process and coupling them to the first n particles of the interacting particle process as described above, we can show that the result holds for each of the n particles. Since each copy of the McKean-Vlasov process is independent of the other copies, we also obtain that the first n particles of the interacting particle system become *asymptotically independent*. This result is what is called the *propagation of chaos* property of the system. The chaos (independence) of the initial condition propagates forward to all the marginals in the limit of the amount of particles N going to infinity.

We also consider a different limit. Instead of sending N to infinity, we keep N fixed and consider the limit $\varepsilon \to 0$, the so called *small noise regime*. If we directly set $\varepsilon = 0$ in Equation (1.1) and Equation (1.2), we obtain two ordinary differential equations (ODEs). The ODE associated to the McKean-Vlasov process is given by

$$\begin{aligned} \frac{\mathrm{d}\varphi(t)}{\mathrm{d}t} &= f(\varphi(t)) + g(\varphi(t-\delta)), \quad t \in [0,T], \\ \varphi(s) &= x(s), \quad s \in [-\delta, 0], \end{aligned} \tag{1.3}$$

for some (deterministic) function $x(s)$. For the interacting particle system we consider two cases. If we assume that the initial conditions are the same for all particles φ^i, we then have

$$\frac{1}{N}\sum_{j=1}^{N} \varphi^j(t) = \varphi^1(t)$$

by symmetry of the equations and the fact that the evolution of the equations is now deterministic. Therefore we find that Equation (1.3) is also the ODE associated to the interacting particle model. In the case that the initial conditions are different we have to use a different approach. This is described in Section 5.3.2.

We are now interested whether or not the stochastic processes converge towards the solutions of these ODEs when $\varepsilon \to 0$. In the case that they do, we are also interested how far they deviate from the deterministic trajectory when ε is small. In particular, we wish to obtain a *large deviation principle (LDP)*. LDPs describe the probability of rare events on an exponential scale. They do so by describing a rate function I. In Chapter 5 we first present various known theorems regarding LDPs for solutions of SDEs. We then show how these theorems can be applied to the processes we consider in this thesis. For the McKean-Vlasov equation we obtain a LDP with a rate function I given by

$$I_x(\phi) = \frac{1}{2}\int_0^T |\dot{\phi}(t) - f(\phi(t)) - g(\varphi(t-\delta))|^2 \,\mathrm{d}t$$

for $\phi - x \in \mathcal{H}_1([0,T], \mathbb{R}^d)$ and $I_x(\phi) = +\infty$ else. Here φ is the solution to the associated ODE. See Theorem 5.12 for details. We also obtain a similar result for the interacting particle process by considering it as one vectorized equation. See Theorem 5.13 for details on this result.

Furthermore, we use simulations to numerically verify these results and analyse a particular scenario for the associated ODE in Chapter 6. We will be doing this with the *Euler-Maruyama approximation*. In this method, similar to the Euler (forward) method for ODEs, the time interval $[0,T]$ is partitioned into equally sized intervals, each with width h. The approximations of the processes are then computed iteratively over the end points of these intervals. If we then interpolate these approximations appropriately, we can show that this method is consistent. That is, we show that in the limit of $h \to 0$ the expected distance between the approximation and the process itself in the supremum norm goes to 0. The exact statement is formulated in Theorem 6.2.

Knowing that the approximation method is consistent, we use simulation to see how certain theoretical results behave in practice. In Section 6.2.1 we verify the results we obtained in Chapter 4 regarding the propagation of chaos property. We find that in the scenario we consider the result indeed holds, and we conjecture that the rate of convergence might be faster than linear. The small noise regime is studied in Section 6.2.2. In particular, we investigate how the McKean-Vlasov processes convergences to the solution of the associated ODE. The results suggest a linear decay, which would be an improvement of the results obtained in Chapter 5. Lastly, we show the effect the delay term can introduce in Section 6.2.3. We show that, for certain choices for the function g, the associated ODE has periodic solutions, while the solution convergences to a constant when we remove the delay.

However, before we can show any of the above, we first have to show that these statements are meaningful, by showing that there exist solutions to Equation 1.1 and Equation 1.2. In Chapter 3 we show that under certain assumptions there exist indeed unique solutions to these equations. We assume that g is globally Lipschitz and that f is locally Lipschitz and satisfies a one-sided Lipschitz assumption. In particular, we do not assume that f satisfies some sub-linear growth condition. This allows us to consider functions of higher orders, given that they still satisfy the one-sided Lipschitz assumption. See Theorem 3.1 and Theorem 3.3 for details for the interacting particle process and the McKean-Vlasov process respectively. We also give assumptions under which we can show that for all $p > 0$, the p'th moment of the process in the supremum norm exists. For the McKean-Vlasov process we do not need to make any further assumptions to show this, see Theorem 3.4. For the interacting particle system we make the additional assumption that g is bounded, see Theorem 3.2. These moment bounds also play an important role in the proofs of the theorems of the later chapters.

In order to make the thesis self-contained, we start with a discussion of the general theory of SDEs in Chapter 2. We will first heuristically derive the meaning behind the stochastic Itô integral. We then formalise these ideas and derive various properties of the Itô integral. In particular, we make great use of the fact that the Itô integral is a martingale. Next we define what we call a solution of a SDE, and present various theorems regarding the existence and uniqueness of solutions in the standard case.

2. Stochastic calculus

Before we are able to analyse our model, we first need to discuss some preliminary theory about SDEs. In particular, we have to define what we mean exactly with a SDE and find ways to do computations with them. We will be doing this in a couple of steps. Throughout this chapter we will follow Chapter 1 and 2 of Mao (2007). The theorems presented here are taken from these chapters unless specified otherwise.

We first explain the idea behind SDEs from a modeling perspective and derive a heuristic definition for a stochastic integral based on what properties we desire it to have. We then find a way to precisely define the stochastic integral. Based on this definition we are able to derive a series of equalities involving the stochastic integral. Furthermore, we find that under mild assumptions on the integrand, the stochastic integral becomes a martingale. Next we present various theorems about martingales, which will be useful in the context of SDEs.

Finally we will define what we call a SDE and in particular define what it means for a process to be a solution. We present conditions under which these solutions exists and we also present Itô's lemma, the stochastic equivalent of the chain rule of calculus. This lemma plays a central role in stochastic calculus and is used in a lot of proofs of theorems about SDEs.

2.1. Stochastic integral

Consider the following ODE, describing the evolution of a process $\{X(t)\}_{0\leq t\leq T}$ taking values in $\mathbb{R}^d$ over a time interval $[0,T]$

$$\begin{aligned} \frac{\mathrm{d}X(t)}{\mathrm{d}t} &= b(X(t),t), \quad t\in[0,T], \\ X(0) &= x_0, \end{aligned}$$

with $b(x,t):\mathbb{R}^d\times[0,T]\to\mathbb{R}^d$ and $x_0\in\mathbb{R}^d$. Equivalently, we can write this problem in an integral form, namely

$$X(t) = x_0 + \int_0^t b(X(t),t)\,\mathrm{d}t, \quad t\in[0,T].$$

Now suppose that the evolution of the process $\{X(t)\}$ is not deterministic, but rather subject to some outside noise. We would wish to model this process with an equation of the form

$$\begin{aligned} \frac{\mathrm{d}X(t)}{\mathrm{d}t} &= b(X(t),t) + \sigma(X(t),t)\cdot \mathit{noise}, \quad t\in[0,T], \\ X(0) &= x_0, \end{aligned}$$

or equivalently

$$X(t) = x_0 + \int_0^t b(X(t),t)\,\mathrm{d}t + \int_0^t \sigma(X(t),t)\cdot \mathit{noise}\,\mathrm{d}t, \quad t\in[0,T], \tag{2.1}$$

with $\sigma(x,t):\mathbb{R}^d\times[0,T]\to\mathbb{R}^{d\times m}$, where m is the dimension of the *noise* vector. Typically we use $m=d$ such that σ is a square matrix, but this is not required.

We would wish that this *noise* process possessed certain properties. Namely we would want that:

1. The *noise* process is unbiased, meaning that its mean is 0;

2. The *noise* at time t is independent of the *noise* at time s for $t \neq s$;

3. The variance of the *noise* is constant over time.

Let us now consider a process that captures only the accumulation of the *noise* in a one-dimensional setup. Suppose that $d = m = 1$ and that $\{X(t)\}$ satisfies the equation

$$X(t) = \int_0^t \mathit{noise}\,\mathrm{d}t, \quad t \in [0, T].$$

If we now furthermore assume that the marginal distributions of $\{X(t)\}$ are normal, then $\{X(t)\}$ becomes a Brownian motion. A Brownian motion is defined as follows.

Definition 2.1 (Brownian motion; 1.4.1 in Mao (2007))**.** *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with filtration $\{\mathcal{F}_t\}_{t\geq 0}$. A (standard) one-dimensional Brownian motion is a real-valued continuous $\{\mathcal{F}_t\}$-adapted process $\{B(t)\}_{t\geq 0}$ with the following properties:*

1. *$B(0) = 0$ a.s.;*

2. *for $0 \leq s < t < \infty$, the increment $B(t) - B(s)$ is normally distributed with mean zero and variance $t - s$;*

3. *for $0 \leq s < t < \infty$, the increment $B(t) - B(s)$ is independent of $\mathcal{F}_s$.*

Remark. *Instead of working with a given filtration, we can also consider a Brownian motion on the* natural filtration *$\{\mathcal{F}_t^B\}_{t\geq 0}$, the filtration generated by the Brownian motion. In this case, property 3 now requires that the process has independent increments.*

We also define a higher dimensional Brownian motion as follows.

Definition 2.2 (1.4.3 in Mao (2007))**.** *A d-dimensional process $\{B(t) = (B^1(t), \ldots, B^d(t))\}_t$ is called a d-dimensional process if every $\{B^i(t)\}_t$ is a one-dimensional Brownian motion, and $\{B^1(t)\}, \ldots, \{B^d(t)\}$ are independent.*

Based on this we would like to define the *noise* process as the derivative of a Brownian motion $B(t)$, such that $\mathit{noise}\,\mathrm{d}t = \mathrm{d}B(t)$. We, however, have that a Brownian motion is nowhere differentiable almost surely. Therefore we would need a new definition for the integral

$$I(t) = \int_0^t f(t)\,\mathrm{d}B(t). \tag{2.2}$$

In order to define this integral, a similar approach is used as in the definition of the normal Riemann integral. That is, we first define the integral for a set of simple functions and then define the general integral as a limit of integrals along these simple functions. We wish to define the integral (2.2) for all $f \in \mathcal{M}^2([a, b]; \mathbb{R}^d)$. This space is defined as follows.

Definition 2.3 (1.5.1 in Mao (2007))**.** *Let $0 \leq a < b < \infty, d \geq 1$. Denote by $\mathcal{M}^p([a, b], \mathbb{R}^d)$ the space of all $\mathbb{R}^d$-valued measurable $\{\mathcal{F}_t\}$-adapted processes $f = \{f(t)\}_{a\leq t\leq b}$ such that*

$$\|f\|_{a,b}^p = \mathbb{E}\left[\int_a^b |f(t)|^p\,\mathrm{d}t\right] < \infty.$$

We say that f and g are equivalent if $\|f - g\|_{a,b}^p = 0$.

We also define the following space of processes which belong to L^p a.s.

Definition 2.4. *Let $0 \leq a < b < \infty, d \geq 1$. Denote by $\mathcal{L}^p([a,b],\mathbb{R}^d)$ the space of all $\mathbb{R}^d$-valued measurable $\{\mathcal{F}_t\}$-adapted processes $f = \{f(t)\}_t$ such that*

$$\int_a^b |f(t)|^p \,\mathrm{d}t < \infty \quad a.s.$$

As said, before defining the integral on the entirety of $\mathcal{M}^2([a,b],\mathbb{R}^d)$, we first define it on a smaller family of processes instead, namely the family of *simple processes.* These are defined as follows.

Definition 2.5 (1.5.2 in Mao (2007))**.** *A $\mathbb{R}^d$-valued process $g = \|g(t)\|_{a\leq t\leq b}$ is called a simple (or step) process if there exists a partition $a = t_0 < t_1 < \cdots < t_k = b$ of $[a,b]$ and bounded random variables $\xi_i, 0 \leq i \leq k-1$ such that ξ_i is $\mathcal{F}_{t_i}$ measurable and*

$$g(t) = \xi_0 \mathbb{1}\{t \in [t_0,t_1]\} + \sum_{i=1}^{k-1} \xi_i \mathbb{1}\{t \in (t_i,t_{i+1}]\} \quad a.s. \tag{2.3}$$

Denote by $\mathcal{M}_0([a,b];\mathbb{R}^d)$ the family of all such processes.

We can now define the stochastic integral for simple processes.

Definition 2.6 (Itô's integral; 1.5.3 in Mao (2007))**.** *For a simple process $g \in \mathcal{M}_0([a,b];\mathbb{R}^d)$ with the form of* (2.3) *define*

$$\int_a^b g(t)\,\mathrm{d}B(t) = \sum_{i=0}^{k-1} \xi_i (B(t_{i+1}) - B(t_i))$$

and call it the Itô integral of g with respect to the Brownian motion $\{B(t)\}$.

Notice that in this definition we essentially evaluate the process g on the left-end point of each interval. While for the normal Riemann integral it does not matter on which end-point we evaluate the integral, it does matter for the stochastic integral. When we use the left-end point we obtain the Itô integral as mentioned above. We will be using this integral throughout the thesis. When we instead use the mean of the left- and the right-end point we obtain the Stratonovich integral, which possesses different properties, which we will not discuss further here. We would also like to remark that the Itô integral is not monotonic. That is, if $f \geq g$, we do not have that $\int_a^b f(t)\,\mathrm{d}B(t) \geq \int_a^b g(t)\,\mathrm{d}B(t)$. This is due to the fact that the terms $B(t_{i+1} - B(t_i)), i = 0, \ldots, k-1$ can also be negative.

To extend this definition to the entirety of $\mathcal{M}^2([a,b];\mathbb{R}^d)$, we present the following approximation lemma.

Lemma 2.7 (1.5.6 in Mao (2007))**.** *For any $f \in \mathcal{M}^2([a,b];\mathbb{R}^d)$, there exists a sequence $\{g_n\}$ of simple processes such that*

$$\lim_{n\to\infty} \mathbb{E}\left[\int_a^b |f(t) - g_n(t)|^2\right] \mathrm{d}t = 0.$$

We are now ready to define the integral (2.2) for all $f \in \mathcal{M}^2([a,b];\mathbb{R}^d)$ as follows.

Definition 2.8 (Itô's integral continued; 1.5.7 in Mao (2007))**.** *Let $f \in \mathcal{M}^2([a,b];\mathbb{R}^d)$. The Itô integral of f with respect to $\{B(t)\}$ is defined by*

$$\int_a^b f(t)\,\mathrm{d}B(t) = \lim_{n\to\infty} \int_a^b g_n(t)\,\mathrm{d}B(t), \tag{2.4}$$

where $(g_n)_n$ is a sequence of simple processes such that

$$\lim_{n\to\infty} \mathbb{E}\left[\int_a^b |f(t) - g_n(t)|^2 \,\mathrm{d}t\right] = 0.$$

Using the previous lemma we find that such a sequence of simple processes always exists. Furthermore, it can also easily be shown that the limit in the right hand side of (2.4) exists and is independent of the particular sequence $\{g_n\}$. Therefore we have that the Itô integral is well defined for all processes $f \in \mathcal{M}^2([a,b];\mathbb{R}^d)$.

Next we state some important properties of the Itô integral.

Theorem 2.9 (1.5.8 and 1.5.9 in Mao (2007)). *Let $f, g \in \mathcal{M}^2([a,b];\mathbb{R}^d)$, and let α, β be two real numbers. Then*

1. $\int_a^b f(t)\,\mathrm{d}B(t)$ *is $\mathcal{F}_b$-measurable;*
2. $\mathbb{E}\left[\int_a^b f(t)\,\mathrm{d}B(t) \mid \mathcal{F}_a\right] = 0$*;*
3. $\mathbb{E}\left[\left|\int_a^b f(t)\,\mathrm{d}B(t)\right|^2\right] = \mathbb{E}\left[\int_a^b |f(t)|^2\,\mathrm{d}t\right]$ *(Itô's isometry);*
4. $\int_a^b \alpha f(t) + \beta g(t)\,\mathrm{d}B(t) = \alpha \int_a^b f(t)\,\mathrm{d}B(t) + \beta \int_a^b g(t)\,\mathrm{d}B(t)$.

These statements are proven by first showing them for simple processes based on definition 2.6 and then showing that they can be extended to all processes in $\mathcal{M}^2([0,T],\mathbb{R}^d)$. Based on this theorem, and in particular point two we get the impression that the stochastic integral is a martingale, as the increments have expectation 0. The following theorem shows that this is indeed the case.

Theorem 2.10 (1.5.14 in Mao (2007)). *Let $T > 0$. Let $f \in \mathcal{M}^2([0,T];\mathbb{R}^d)$. Define*

$$I(t) = \int_0^t f(s)\,\mathrm{d}B(s), \quad 0 \le t \le T,$$

with $I(0) = 0$. We call $I(t)$ the indefinite Itô integral of f. We now have that $I = \{I(t)\}_{0 \le t \le T}$ is a square integrable continuous martingale with quadratic variation given by

$$\langle I, I\rangle_t = \int_0^t |f(s)|^2\,\mathrm{d}s, \quad 0 \le t \le T.$$

Here the quadratic variation is defined as the unique process starting at 0 such that $I^2 - \langle I, I\rangle$ is a martingale. Since I is continuous we also have the equivalent definition

$$\langle I, I\rangle_t = [I]_t = \lim_{\|P\| \to 0} \sum_{k=1}^n (I(t_k) - I(t_{k-1}))^2,$$

where P ranges over the partitions of the interval $[0,t]$.

Remark. *If we use $f = 1$, we obtain that $I(t) = B(t)$ based on Definition 2.6. From this we can conclude that the quadratic variation $[B]_t = t$. In particular we have that*

$$\lim_{\|P\| \to 0} \sum_{k=1}^n (B(t_k) - B(t_{k-1}))^2 = t = \lim_{\|P\| \to 0} \sum_{k=1}^n (t_k - t_{k-1}),$$

which suggests that, heuristically speaking, $\mathrm{d}B(t)^2 = \mathrm{d}t$.

Remark. *It is also possible to extend the definition of the Itô integral even further to the space $\mathcal{L}^2([a,b],\mathbb{R}^d)$. In this case the integral I is no longer a martingale but only a local martingale. This allows us to use the Itô integral before we verify that the integrand belongs to $\mathcal{M}^2([a,b],\mathbb{R}^d)$. We can than later check that this is indeed the case, such that the integral still becomes a martingale.*

2.2. Martingale inequalities

Knowing that the stochastic integral is a martingale, we now present some theorems about martingale processes that will be useful later.

We first present Doob's inequality. This inequality relates the moments of the supremum over a time interval $[a, b]$ of a martingale process to the moments of that martingale at time b. As we wish to bound the supremum norm of certain processes, or the differences between them, this seems like a useful inequality.

Theorem 2.11 (Doob's inequality; 1.3.8 in Mao (2007))**.** *Let $\{M(t)\}$ be an $\mathbb{R}^d$-valued martingale. Let $[a, b]$ be a bounded interval of $\mathbb{R}_+$. Let $p > 1$ and suppose that for all t, $M(t) \in L^p$, i.e. $\mathbb{E}[|M(t)|^p] < \infty$. Then*

$$\mathbb{E}\left[\sup_{a \le t \le b} |M_t|^p\right] \le \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[|M(b)|^p\right]. \tag{2.5}$$

The downside of Doob's inequality is that we still need to manage the moments of martingale at time b. In the case that the martingale is a stochastic integral, that means we want to compute

$$\mathbb{E}\left[\left|\int_a^b f(s)\,\mathrm{d}B(s)\right|^p\right].$$

In general, it is not directly clear how to do this. Point 3 of Theorem 2.9 however shows that for $p = 2$ there is a relation between the second moment of the stochastic integral and the L^2 norm of the integrand. Let $f \in \mathcal{M}^2([0, T], \mathbb{R}^{d\times m})$ and set $I(t) = \int_0^t f(s)\,\mathrm{d}B(s)$. Using Doob's inequality and Itô's isometry we have

$$\mathbb{E}\left[\sup_{0 \le t \le T} |I(t)|^p\right] \le 4\mathbb{E}\left[\int_0^T |f(s)|^2\,\mathrm{d}s\right].$$

It turns out that this inequality can be generalised for all $p \ge 2$, as the following theorem shows.

Theorem 2.12 (1.7.2 in Mao (2007))**.** *Let $p \ge 2$. Let $\{B(t)\}$ be a m-dimensional Brownian motion and let $g \in \mathcal{M}^2([0, T], \mathbb{R}^{d\times m})$ such that*

$$\mathbb{E}\left[\int_0^T |g(s)|^p\,\mathrm{d}s\right] < \infty.$$

Then

$$\mathbb{E}\left[\sup_{0 \le t \le T}\left|\int_0^t g(s)\,\mathrm{d}B(s)\right|^p\right] \le \left(\frac{p^3}{2(p-1)}\right)^{\frac{p}{2}} T^{\frac{p-2}{2}}\mathbb{E}\left[\int_0^T |g(s)|^p\,\mathrm{d}s\right].$$

This inequality turns out to be really useful, as it not only gets rid of the supremum, but also changes the stochastic integral into a normal integral and shifts the power p inside the integral. We will be using this theorem in the proofs of various theorems, such as the existence to solutions of stochastic differential equations. We would also like to compare it to the following lemma, which also allows us to move the power p inside the integral, but this time for normal integrals.

Lemma 2.13. *Let $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Let $f \in \mathcal{M}^p([0, T], \mathbb{R}^d)$, then*

$$\mathbb{E}\left[\sup_{0 \le t \le T}\left|\int_0^t f(s)\,\mathrm{d}s\right|^p\right] \le T^{p-1}\mathbb{E}\left[\int_0^T |f(s)|^p\,\mathrm{d}s\right].$$

Proof. Firstly, we have

$$\sup_{0\le t\le T}\left|\int_0^t f(s)\,\mathrm{d}s\right|^p \le \left(\int_0^T |f(s)|\,\mathrm{d}s\right)^p.$$

Using Hölders theorem we have

$$\begin{aligned}\left(\int_0^T |f(s)|\,\mathrm{d}s\right)^p &\le \left(\int_0^T 1^q\,\mathrm{d}s\right)^{\frac{p}{q}} \int_0^T |f(s)|^p\,\mathrm{d}s\\ &= T^{\frac{p}{q}}\int_0^T |f(s)|^p\,\mathrm{d}s.\end{aligned}$$

Using that $\frac{1}{p}+\frac{1}{q}=1$ we can deduce that $\frac{p}{q}=p-1$. Combining these properties and taking expectation on both sides completes the proof. ■

The only downside to Theorem 2.12 is that the statement is only valid for $p\ge 2$. In certain cases we also would like to be able to control for example the first moment. For this we present the following theorem, that will help us out in those scenarios.

Theorem 2.14 (Burkholder-Davis-Gundy Inequality)**.** *Let $\{X(t)\}$ be a martingale with $X(0)=0$ and quadratic variation $\{A(t)\}$. Then for every $p>0$, there exist a universal positive constant c_p, C_p (depending only on p), such that*

$$c_p\mathbb{E}\left[|A(t)|^{\frac{p}{2}}\right] \le \mathbb{E}\left[\sup_{0\le s\le t}|X(s)|^p\right] \le C_p\mathbb{E}\left[|A(t)|^{\frac{p}{2}}\right].$$

In particular, in the case that $\{X(t)\}$ is giving by

$$X(t)=\int_0^t f(s)\,\mathrm{d}B(s),\quad t\in[0,T],$$

$f\in\mathcal{M}^2([0,T],\mathbb{R}^d)$, we have

$$c_p\mathbb{E}\left[\left|\int_0^t |f(s)|^2\,\mathrm{d}s\right|^{\frac{p}{2}}\right] \le \mathbb{E}\left[\sup_{0\le s\le t}|\int_0^t f(s)\,\mathrm{d}B(s)|^p\right] \le C_p\mathbb{E}\left[\left|\int_0^t |f(s)|^2\,\mathrm{d}s\right|^P\right].$$

2.3. Stochastic differential equations (SDEs)

Now that we have defined the stochastic integral, we can return to our original problem. Equation (2.1) can now be written as

$$X(t)=x_0+\int_0^t b(X(t),t)\,\mathrm{d}t+\int_0^t \sigma(X(t),t)\,\mathrm{d}B(t),\quad t\in[0,T]. \tag{2.6}$$

We now wish to seek solutions $\{X(t)\}$ that satisfy $\{\sigma(X(t),t)\}\in\mathcal{M}^2([0,T];\mathbb{R}^d)$ such that the stochastic integral is well-defined. Before we are able to continue, we first need an important result to more properly deal with the stochastic integral. Similar to the normal Riemann integrals, the definition as is is not particularly helpful when evaluating the integral. Itô's lemma (also known as Itô's formula) will help us with this. We first define what am Itô process is

Definition 2.15 (1.6.3 in Mao (2007))**.** *A d-dimensional Itô process is an $\mathbb{R}^d$-valued continuous adapted process $\{X(t)\} = \{(X_1(t), \ldots, X_d(t))^T\}$ on $t \geq 0$ of the form*

$$X(t) = X(0) + \int_0^t f(s)\,\mathrm{d}s + \int_0^t g(s)\,\mathrm{d}B(s),$$

where $f = (f_1, \ldots, f_d)^T \in \mathcal{L}^1(\mathbb{R}_+; \mathbb{R}^d)$,$g = (g_{ij})_{d\times m} \in \mathcal{L}^2(\mathbb{R}_+; \mathbb{R}^{d\times m})$, and $B(t)$ is a m-dimensional Brownian motion. We say that $\{X(t)\}$ has stochastic differential $\mathrm{d}X(t)$ given by

$$\mathrm{d}X(t) = f(t)\mathrm{d}t + g(t)\mathrm{d}B(t).$$

Notice that if we where to found solutions to (2.6), that this process would be an Itô process with $f(t) = b(X(t), t)$ and $g(t) = \sigma(X(t), t)$, given that they are integrable and square integrable respectively.

We now have the follow theorem.

Theorem 2.16 (Itô's lemma; 1.6.4 in Mao (2007))**.** *Let $\{X(t)\}$ be a d-dimensional Itô process on $t \geq 0$ with stochastic differential*

$$\mathrm{d}X(t) = f(t)\mathrm{d}t + g(t)\mathrm{d}B(t),$$

with $f \in \mathcal{L}^1(\mathbb{R}_+; \mathbb{R}^d)$ and $g \in \mathcal{L}^2(\mathbb{R}_+; \mathbb{R}^{d\times m})$. Let $V \in C^{2,1}(\mathbb{R}^d \times \mathbb{R}_+; \mathbb{R})$. Then $\{V(X(t), t)\}$ is again an Itô process with stochastic differential given by

$$\begin{aligned}\mathrm{d}V(X(t), t) = &\left[\frac{\partial V(X(t), t)}{\partial t} + (\nabla_x V(X(t), t))^T f(t) + \frac{1}{2}\mathrm{trace}\left(g(t)^T H_x(V(X(t), t))g(t)\right)\right]\mathrm{d}t \\ &+ (\nabla_x V(X(t), t))^T g(t)\mathrm{d}B(t),\end{aligned}$$

where ∇_x denotes the gradient with respect to the x variable and H_x the Hessian with respect to the x variable.

Notice how this result is very similar to the usual chain rule in calculus. The only difference is that in this case the second derivative still plays a roll. We will show why this is the case by giving a heuristic idea of the proof for the case $d = 1$. We will consider a Taylor expansion of $V(x, t)$. We obtain that

$$\mathrm{d}V(x, t) = \frac{\partial V(x, t)}{\partial t}\mathrm{d}t + \frac{\partial V(x, t)}{\partial x}\mathrm{d}x + \frac{1}{2}\frac{\partial^2 V(x, t)}{\partial x^2}\mathrm{d}x^2 + \cdots.$$

Setting $x = X(t)$ and using the expression for $\mathrm{d}X(t)$ we get

$$\begin{aligned}\mathrm{d}V(X(t), t) = &\frac{\partial V(X(t), t)}{\partial t}\mathrm{d}t + \frac{\partial V(X(t), t)}{\partial x}\left(f(t)\mathrm{d}t + g(t)\mathrm{d}B(t)\right) \\ &+ \frac{1}{2}\frac{\partial^2 V(X(t), t)}{\partial x^2}\left(f(t)^2\mathrm{d}t^2 + 2f(t)g(t)\mathrm{d}t\mathrm{d}B(t) + g(t)^2\mathrm{d}B(t)^2\right) + \cdots.\end{aligned}$$

The quadratic variation of the Brownian motion now suggests that $\mathrm{d}B(t)^2 = \mathrm{d}t$, while the terms $\mathrm{d}t^2$, $\mathrm{d}t\mathrm{d}B(t)$ and all higher order terms tend to 0 faster than $\mathrm{d}t$. Therefore we obtain, heuristically,

$$\begin{aligned}\mathrm{d}V(X(t), t) &= \frac{\partial V(X(t), t)}{\partial t}\mathrm{d}t + \frac{\partial V(X(t), t)}{\partial x}\left(f(t)\mathrm{d}t + g(t)\mathrm{d}B(t)\right) + \frac{1}{2}\frac{\partial^2 V(X(t), t)}{\partial x^2}g(t)^2\mathrm{d}t \\ &= \left(\frac{\partial V(X(t), t)}{\partial t} + \frac{\partial V(X(t), t)}{\partial x}f(t) + \frac{1}{2}\frac{\partial^2 V(X(t), t)}{\partial x^2}g(t)^2\right)\mathrm{d}t \\ &\quad + \frac{\partial V(X(t), t)}{\partial x}g(t)\mathrm{d}B(t).\end{aligned}$$

We now have all the tools we need to find solutions to Equation (2.6). First, we define what we mean with a solution.

Definition 2.17 (2.2.1 in Mao (2007)). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space with filtration $\{\mathcal{F}(t)\}_t$. Let $\{B(t)\}$ be a m-dimensional Brownian motion on this probability space. An $\mathbb{R}^d$-valued stochastic process $\{X(t)\}$ is called a solution of Equation (2.6) if it has the following properties:*

1. *$\{X(t)\}$ is continuous and $\mathcal{F}_t$-adapted;*
2. *$\{f(X(t),t)\} \in \mathcal{L}^1([0,T];\mathbb{R}^d)$ and $\{g(X(t),t)\} \in \mathcal{L}^2([0,T],\mathbb{R}^{d\times m})$;*
3. *Equation (2.6) holds for every $t \in [0,T]$ with probability 1.*

In this definition that we will use from now on, the Brownian motion $\{B(t)\}$ is assumed to be part of the equation, rather than the solution, and we say that Equation (2.6) has a strong solution if it has a solution for all Brownian motions. There exists also the notion of weak solutions for (2.6) in which the Brownian motion and the probability space itself are instead a part of the solution, but we will not discuss this here further. We now have the following important theorem regarding the existence of strong solutions for Equation (2.6).

Theorem 2.18 (2.3.1 in Mao (2007)). *Consider the SDE (2.6). Assume that $\mathbb{E}\left[|x_0|^2\right] < \infty$. Assume that the function $b(x,t)$ and $\sigma(x,t)$ are Lipschitz in space, i.e. there exists a $K > 0$ such that for all $x, y \in \mathbb{R}^d, t \in [0,T]$ we have*

$$|b(x,t) - b(y,t)|^2 \vee |\sigma(x,t) - \sigma(y,t)|^2 \leq K|x-y|^2.$$

Furthermore, assume a linear growth condition, i.e. there exists $\tilde{K} > 0$ such that for all $x \in \mathbb{R}^d, t \in [0,T]$ we have

$$|b(x,t)|^2 \vee |\sigma(x,t)|^2 \leq \tilde{K}(1+|x|^2).$$

Under these assumptions we have that there exists a unique solution $\{X(t)\}$ of (2.6) belonging to $\mathcal{M}^2([0,T],\mathbb{R}^d)$.

Proof. We start the proof by showing the boundedness of the solution. In particular, suppose that $\{X(t)\}$ is a solution of (2.6), we will show that

$$\mathbb{E}\left[\sup_{0\leq s\leq T} |X(s)|^2\right] \leq \left(1 + 3\mathbb{E}\left[|x_0|^2\right]\right)\exp\left(3KT(T+4)\right).$$

Firstly, for every integer $n \geq 1$, define the stopping time $\tau_n = T \wedge \inf\{t \in [0,T] : |X(t)| \geq n\}$. Define $X^n(t) = X(t \wedge \tau_n)$ for $t \in [0,T]$. Then $X^n(t)$ satisfies the equation

$$X^n(t) = x_0 + \int_0^t b(X^n(s),s)\mathbb{1}\{s \in [0,\tau_n]\}\,\mathrm{d}s + \int_0^t \sigma(X^n(s),s)\mathbb{1}\{s \in [0,\tau_n]\}\,\mathrm{d}B(s).$$

Using the binomial inequality gives

$$|X^n(t)|^2 \leq 3|x_0|^2 + \left|\int_0^t b(X^n(s),s)\mathbb{1}\{s \in [0,\tau_n]\}\,\mathrm{d}s\right|^2 + \left|\int_0^t \sigma(X^n(s),s)\mathbb{1}\{s \in [0,\tau_n]\}\,\mathrm{d}B(s)\right|^2.$$

Applying Hölders inequality (Lemma 2.13), the Doob-like inequality (Theorem 2.12) and the growth assumption gives

$$\mathbb{E}\left[\sup_{0\leq s\leq t}|X^n(s)|^2\right] \leq 3\mathbb{E}\left[|x_0|^2\right] + 3\tilde{K}T\int_0^t \left(1+\mathbb{E}\left[|X^n(s)|^2\right]\right)\mathrm{d}s + 12\tilde{K}\int_0^t\left(1+\mathbb{E}\left[|X^n(s)|^2\right]\right)\mathrm{d}s.$$

We therefore have

$$1 + \mathbb{E}\left[\sup_{0\le s\le t} |X^n(s)|^2\right] \le 1 + 3\mathbb{E}\left[|x_0|^2\right] + 3\tilde{K}(T+4)\int_0^t \left(1 + \mathbb{E}\left[|X^n(s)|^2\right]\right) \mathrm{d}s.$$

Applying Gronwall's inequality gives

$$1 + \mathbb{E}\left[\sup_{0\le s\le T} |X^n(s)|^2\right] \le \left(1 + 3\mathbb{E}\left[|x_0|^2\right]\right) \exp\left(3KT(T+4)\right).$$

Using the definition of $X^n(t)$, we have that

$$n^2\mathbb{P}(\tau_n < T) \le \mathbb{E}\left[\sup_{0\le s\le T} |X^n(s)|^2\right].$$

Since we just showed that the right hand side is bounded independent of n, we clearly have $\mathbb{P}(\tau_n < T) \to 0$ as $n \to \infty$. Therefore we have that $X^n \to X$ almost surely. Applying Fatou's lemma now proves the required bound.

Next we prove uniqueness. Suppose that $\{X(t)\}$ and $\{Y(t)\}$ are two solutions of (2.6). We obtain that

$$X(t) - Y(t) = \int_0^t b(X(s), s) - b(Y(s), s)\,\mathrm{d}s + \int_0^t \sigma(X(s), s) - \sigma(Y(s), s)\,\mathrm{d}B(s),$$

$$|X(t) - Y(t)|^2 \le 2\left|\int_0^t b(X(s), s) - b(Y(s), s)\,\mathrm{d}s\right|^2 + 2\left|\int_0^t \sigma(X(s), s) - \sigma(Y(s), s)\,\mathrm{d}B(s)\right|^2.$$

Taking supremum and expectation, and then using Hölders inequality (Lemma 2.13) and the Doob-like inequality (Theorem 2.12) we obtain

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le r\le t} |X(t) - Y(t)|^2\right] \le{}& 2t\mathbb{E}\left[\int_0^t |b(X(s), s) - b(Y(s), s)|^2\,\mathrm{d}s\right] \\ &+ 8\mathbb{E}\left[\int_0^t |\sigma(X(s), s) - \sigma(Y(s), s)|^2\,\mathrm{d}s\right].\end{aligned}$$

Using the Lipschitz assumption we get

$$\mathbb{E}\left[\sup_{0\le r\le t} |X(t) - Y(t)|^2\right] \le K(2T+8)\int_0^t \mathbb{E}\left[\sup_{0\le r\le s} |X(s) - Y(s)|^2\right] \mathrm{d}s.$$

Finally, applying Grönwall's inequality gives

$$\mathbb{E}\left[\sup_{0\le r\le t} |X(t) - Y(t)|^2\right] = 0.$$

Therefore we conclude that $\{X(t)\}$ and $\{Y(t)\}$ are indistinguishable, proving the uniqueness.

Lastly, we present the proof for the existence of solutions. For this we make use of Picard iterations. Namely, let $X^0(t) = x_0$ for all $t \in [0, T]$ and for every integer $m \ge 0$ define

$$X^{m+1}(t) = x_0 + \int_0^t b(X^m(s), s)\,\mathrm{d}s + \int_0^t \sigma(X^m(s), s)\,\mathrm{d}B(s).$$

We will show that this is a contraction. In particular, we prove using induction that

$$\mathbb{E}\left[\sup_{0\le r\le t} |X^{m+1}(r) - X^m(r)|^2\right] \le \frac{(Rt)^{m+1}}{(m+1)!},$$

with $R = \max\{2\tilde{K}(T+8)(1+\mathbb{E}[|x_0|^2]), 2(T+4)K\}$. For $m = 0$ we compute, using Hölders inequality,

$$\sup_{0 \le r \le t} |X^1(r) - x_0|^2 \le 2t \int_0^t |b(x_0, s)|^2 \,\mathrm{d}s + \sup_{0 \le r \le t} 2 \left| \int_0^r \sigma(x_0, s) \,\mathrm{d}s \right|^2 .$$

Taking expectation, using the Doob-like inequality and the linear growth assumption gives

$$\begin{aligned}
\mathbb{E}\left[\sup_{0 \le r \le t} |X^1(r) - x_0|^2\right] &\le 2\tilde{K}(T+8) \int_0^t \mathbb{E}\left[1 + |x_0|^2\right] \,\mathrm{d}s \\
&\le 2\tilde{K}(T+8)t(1+\mathbb{E}\left[|x_0|^2\right]) \le Rt.
\end{aligned}$$

Now suppose that the result holds for $m-1$. Now for m we have

$$\begin{aligned}
\sup_{0 \le r \le t} |X^{m+1}(r) - X^m(r)|^2 \le 2 \sup_{0 \le r \le t} &\left| \int_0^r b(X^m(s), s) - b(X^{m-1}(s)) \,\mathrm{d}s \right|^2 \\
&+ 2 \sup_{0 \le r \le t} \left| \int_0^r \sigma(X^m(s), s) - \sigma(X^{m-1}(s)) \,\mathrm{d}B(s) \right|^2 .
\end{aligned}$$

Taking expectation, applying Hölder and the Doob-like inequality and applying the Lipschitz assumption we get

$$\begin{aligned}
\mathbb{E}\left[\sup_{0 \le r \le t} |X^{m+1}(r) - X^m(r)|^2\right] &\le 2(T+4)K \int_0^t \mathbb{E}\left[\sup_{0 \le r \le t} |X^m(r) - X^{m-1}(r)|^2\right] \,\mathrm{d}s \\
&\le R \int_0^t \frac{(Rs)^m}{m!} \,\mathrm{d}s \\
&\le \frac{(Rt)^{m+1}}{(m+1)!},
\end{aligned}$$

which proves the result. Applying Markov's inequality now gives

$$\begin{aligned}
\mathbb{P}\left(\sup_{0 \le t \le T} |X^{m+1}(t) - X^m(t)| > \frac{1}{2^m}\right) &\le 2^{2m} \mathbb{E}\left[\sup_{0 \le t \le T} |X^{m+1}(t) - X^m(t)|^2\right] \\
&\le 2^{2m} \frac{(RT)^{m+1}}{(m+1)!}.
\end{aligned}$$

Since this is summable over m, the Borel-Cantelli lemma gives us that

$$\mathbb{P}\left(\sup_{0 \le t \le T} |X^{m+1}(r) - X^m(r)|^2 > \frac{1}{2^m} \text{ for infinitely many indices } m\right) = 0.$$

Therefore, for almost every $\omega \in \Omega$ we have for sufficiently large m that

$$\sup_{0 \le t \le T} |X^{m+1}(r) - X^m(r)|^2 \le \frac{1}{2^m},$$

such that for fixed ω the series

$$x_0 + \sum_{i=0}^{m-1} (X^{k+1}(t) - X^k(t)) = X^m(t)$$

converges uniformly on $[0, T]$. Now define $X(t) = \lim_{m\to\infty} X^m(t)$, $t \in [0, T]$. We now show that $\{X(t)\}$ is a solution to (2.6). Recall that the $\{X^m(t)\}$ are given by

$$X^{m+1}(t) = x_0 + \int_0^t b(X^m(s), s)\,\mathrm{d}s + \int_0^t \sigma(X^m(s), s)\,\mathrm{d}B(s).$$

We clearly have that the left hand side converges to $X(t)$ uniformly. By the Lipschitz assumption we have that $b(X^m(s), s) \to b(X(s), s)$ and $\sigma(X^m(s), s) \to \sigma(X(s), s)$ uniformly, and therefore also the integrals

$$\int_0^t b(X^m(s), s)\,\mathrm{d}s \to \int_0^t b(X(s), s)\,\mathrm{d}s \quad \text{a.s.}$$

and

$$\int_0^t |\sigma(X^m(s), s) - \sigma(X(s), s)|^2\,\mathrm{d}s \to 0 \quad \text{a.s.}$$

Therefore, using Proposition 7.3 of Baldi (2017), which relates the convergence in probability in L^2 of the integrand to convergence of the Itô integral in probability,

$$\int_0^t \sigma(X^m(s), s)\,\mathrm{d}B(s) \xrightarrow{P} \int_0^t \sigma(X^m(s), s)\,\mathrm{d}B(s).$$

Taking the limit along a sub-sequence turns this convergence in probability into almost sure convergence. Since the other terms already converge a.s., by uniqueness of limits this does not cause any problems. Therefore we have proven that $\{X(t)\}$ is a solution to (2.6), proving the theorem. ■

The assumption as stated in Theorem 2.18 can easily be weakened to allow for more general drift functions b and diffusion functions σ. In particular, we will show that a local Lipschitz assumption suffices to prove existence and uniqueness up to a time τ_R, the stopping time when the process first leaves the ball or radius R centered around 0. In order to then prove general existence and uniqueness on a time interval $[0, T]$, one only needs a sufficient growth condition to show that $\lim_{R\to\infty} \tau_R \wedge T = T$ a.s. In this case the solution of the SDE will be defined based on these stopped processes. The reason that this approach works is because of the following theorem.

Theorem 2.19 (9.3 in Baldi (2017))**.** *Let $b_i, \sigma_i, i = 1, 2$, be measurable functions on $\mathbb{R}^d \times [0, T]$. Let $X_i, i = 1, 2$, be solutions of the SDE*

$$\begin{aligned} \mathrm{d}X_i(t) &= b_i(X_i(t), t)\mathrm{d}t + \sigma_i(X_i(t), t)\mathrm{d}B(t). \\ X_i(0) &= x_0. \end{aligned}$$

Let $D \subset \mathbb{R}^d$ be an open set such that, on $D \times [0, T]$, $b_1 = b_2, \sigma_1 = \sigma_2$ and, for every $x, y \in D, 0 \le t \le T$,

$$|b_i(x, t) - b_i(y, t)| \le L|x - y|, \qquad |\sigma_i(x, t) - \sigma_i(y, t)| \le L|x - y|.$$

Then, if τ_i denotes the exit time of X_i from D,

$$\tau_1 \wedge T = \tau_2 \wedge T \quad \text{a.s.} \quad \text{and} \quad \mathbb{P}\left(X_1(t) = X_2(t) \text{ for every } 0 \le t \le \tau_1 \wedge T\right) = 1.$$

Using the above result we can now state the following theorem, based on Theorem 2.3.4 in Mao (2007) and Theorem 9.4 in Baldi (2017).

Theorem 2.20. *Consider the SDE*

$$\begin{aligned}\mathrm{d}X(t) &= b(X(t),t)\mathrm{d}t + \sigma(X(t),t)\mathrm{d}B(t),\\ X(0) &= x_0.\end{aligned}$$

Assume that $\mathbb{E}\left[|x_0|^2\right] < \infty$ *and that the function* b *and* σ *are locally Lipschitz, i.e. for each* $R > 0$ *there exists an constant* $L_R > 0$ *such that for all* x, y *with* $|x| \wedge |y| < R$ *and* $t \in [0, T]$ *we have*

$$|b(x,t) - b(y,t)| \wedge |\sigma(x,t) - \sigma(y,t)| \leq L_R|x-y|.$$

Lastly, assume that $\sup_{0\leq t\leq T} |b(0,t)| \wedge |\sigma(0,t)| < \infty$*. Then there exists a unique solution* $\{X(t)\}$ *on the interval* $[0, \tau_\infty]$*, with* $\tau_\infty = \lim_{R\to\infty} \tau_R \wedge T$*. Here* τ_R *denotes the exit times from the ball with radius* R *centered around 0.*

Proof. Let $R > 0$ we consider the SDE

$$\begin{aligned}\mathrm{d}X_R(t) &= b_R(X_R(t),t)\mathrm{d}t + \sigma_R(X_R(t),t)\mathrm{d}B(t),\\ X_R(0) &= x_0,\end{aligned}$$

where b_R is such that $b_R = b$ on $B(0,R) \times [0,T]$ and globally Lipschitz, and similarly for σ_R. By the assumption that $\sup_{0\leq t\leq T} |b(0,t)| \wedge |\sigma(0,t)| < \infty$ we have that b_R and σ_R also satisfy the linear growth condition of Theorem 2.18. Therefore we can apply Theorem 2.18 to obtain a unique solution $X_R(t)$. We now define the solution $\{X(t)\}$ of the original process as

$$X(t) = X_R(t), \quad t \leq \tau_R.$$

Because of Theorem 2.19 this solution is well-defined. Indeed let $0 < R_1 < R_2$. Then we have that $b_{R_1} = b = b_{R_2}$ and $\sigma_{R_1} = \sigma = \sigma_{R_2}$ on $B(0,R_1) \times [0,T]$. Therefore $X_{R_1} = X_{R_2}$ on $t < \tau_{R_1}$, and $\tau_{R_2} \geq \tau_{R_1}$ a.s. We now have constructed a solution on the interval $[0, \tau_\infty]$, as required, proving the theorem. ■

Remark. *The equations we have considered so far feature functions* b *and* σ *which on itself are deterministic. It is however also possible to make the functions itself depend on* ω*. It turns out that the above theory is still valid as long as the functions* b *and* σ *are adapted and of course still satisfy the other required assumptions. Details can be found in Gikhman and Skorohod (1972).*

3. Existence, uniqueness and moment bounds

In this chapter we will investigate under which conditions the equations presented in Chapter 1 have strong solutions. In Chapter 2 we have seen sufficient conditions for the existence and uniqueness of basic SDEs.

In this chapter we will relax these assumptions further and show that these milder solutions are sufficient to prove the existence and uniqueness for the class of SDEs with delay we are considering. Furthermore, we show that all the moments of the processes exist, both for the marginals and the supremum norm.

Knowing that the equations possess unique solutions, we can study the solutions and their properties further in the next chapters. In these later chapters the moment bounds also play an important role, as they are used to derive other properties of the solutions.

The chapter is structured as follows; we start by presenting all our results, first for the interacting particle process in Section 3.1 and then for the McKean-Vlasov process in Section 3.2. The proofs for all the results are presented afterwards in Section 3.3.

3.1. Interacting particle process

Let $N \geq 1$ and $T, \delta > 0$. As stated we will first consider the following model, describing an interacting particle system, where the interaction happens through the mean with a delay δ.

$$\mathrm{d}X^i(t) = f(X^i(t))\mathrm{d}t + g\left(\frac{1}{N}\sum_{j=1}^{N} X^j(t-\delta)\right)\mathrm{d}t + \varepsilon \mathrm{d}B^i(t), \quad i \in 1, \ldots, N, t \in [0,T], \tag{3.1}$$

$$X^i(0) = x_0^i, \tag{3.2}$$

$$X^i(s) = \xi_s^i, \quad s \in [-\delta, 0), \tag{3.3}$$

with $f, g \in C(\mathbb{R}^d, \mathbb{R}^d)$ and B^i independent d-dimensional Brownian motions. We assume that x_0^i and $\xi_s^i, s \in [-\delta, 0)$ are i.i.d. and $\mathcal{F}_0$-measurable. For the functions f and g we have the following set of assumptions.

Assumption 1. *We say that the functions f and g satisfy Assumption 1 if the following holds.*

1. *The function f is locally Lipschitz, i.e. for each $R > 0$ there exists a constant $L_R > 0$ such that for all $x, y \in \mathbb{R}^d$ with $|x| \wedge |y| < R$ we have that*
$$|f(x) - f(y)| \leq L_R|x - y|.$$
2. *The function f satisfies a one-sided global Lipschitz condition, i.e. there exists a constant $C > 0$ such that for all $x, y \in \mathbb{R}^d$ we have*
$$\langle x - y, f(x) - f(y)\rangle < C|x - y|^2.$$
3. *The function g is globally Lipschitz, i.e. there exists a constant $L > 0$ such that for all $x, y \in \mathbb{R}^d$ we have*
$$|g(x) - g(y)| \leq L|x - y|.$$

Remark. *Although the above assumptions are sufficient for our result, we make some further minor assumptions to make the proofs simpler. In particular we will assume that $f(0) = g(0) = 0$. This implies that we have the following two inequalities for all $x \in \mathbb{R}^d$:*

$$\langle x, f(x)\rangle \leq C|x|^2,$$
$$|g(x)| \leq L|x|.$$

3.1.1. Existence and uniqueness

For the existence and uniqueness of solutions to Equation (3.1) we have the following theorem.

Theorem 3.1. *Suppose that Assumption 1 holds. Suppose that* $\mathbb{E}\left[|x_0^i|^2\right] < \infty$ *and also that* $\sup_{-\delta \le t < 0} \mathbb{E}\left[|\xi_t^i|^2\right] < \infty$. *Then there exists a unique set of solutions* $\{X^i(t)\}, i = 1, \dots, N$ *to Equation* (3.1). *Furthermore, we have that* $\{X^i(t)\} \in \mathcal{M}^2([0,T], \mathbb{R}^d)$ *for all* $i = 1, \dots, N$.

3.1.2. Moment bounds

In Theorem 3.1 we have seen that the solutions $\{X^i(t)\}$ belong to $\mathcal{M}^2([0,T], \mathbb{R}^d)$. In this section we will extend this results. We have the following theorem.

Theorem 3.2. *Let* f *and* g *be functions that satisfy Assumption 1. Furthermore, assume that* g *is bounded, i.e. there exists a constant* $M > 0$ *such that* $|g(x)| \le M$ *for all* $x \in \mathbb{R}^d$. *Let* $p > 1$. *Assume that for all* $1 \le i \le N$ *we have* $\mathbb{E}\left[|x_0^i|^p\right] < \infty$. *Let* $\{X^i(t)\}, i = 1, \dots, N$ *be the corresponding solutions of Equation* (3.1). *We have*

$$\mathbb{E}\left[\sup_{0 \le t \le T} |X^i(t)|^p\right] < \infty. \tag{3.4}$$

3.2. McKean–Vlasov process

We now instead consider the McKean-Vlasov process, where the dependence on the mean also happens with a delay δ.

$$\begin{aligned} \mathrm{d}X(t) &= f(X(t))\mathrm{d}t + g(\mathbb{E}\left[X(t-\delta)\right])\mathrm{d}t + \varepsilon \mathrm{d}B(t), \quad t \in [0,T], && (3.5)\\ X(0) &= x_0, && (3.6)\\ \mathbb{E}\left[X(s)\right] &= \xi_s, \quad s \in [-\delta, s). && (3.7) \end{aligned}$$

We again have that $f, g \in C(\mathbb{R}^d, \mathbb{R}^d)$ and that B is an d-dimensional Brownian motion. We assume that x_0 is $\mathcal{F}_0$-measurable.

3.2.1. Existence and uniqueness

For the existence and uniqueness of solutions to Equation (3.5) we have the following theorem.

Theorem 3.3. *Suppose that Assumption 1 holds. Suppose that* $\mathbb{E}\left[|x_0|^2\right] < \infty$ *and also that* $\sup_{-\delta \le t < 0} |\xi_t|^2 < \infty$. *Then there exists a unique solution* $\{X(t)\}$ *to Equation* (3.5). *Furthermore, we have that* $\{X(t)\} \in \mathcal{M}^2([0,T], \mathbb{R}^d)$.

3.2.2. Moment bounds

Similar to what we did for the interacting particle process, we now also extend the result that the solution to the McKean-Vlasov process belongs to $\mathcal{M}^2([0,T], \mathbb{R}^d)$. In particular, we have the following theorem.

Theorem 3.4. *Let* $\{X(t)\}$ *be the solution of* (3.5), *where Assumption 1 holds. Let* $p > 1$. *Assume that* $\mathbb{E}\left[|x_0|^p\right] < \infty$. *We have*

$$\mathbb{E}\left[\sup_{0 \le t \le T} |X(t)|^p\right] < \infty. \tag{3.8}$$

3.3. Proofs

We will now present the proof for the above mentioned theorems. We will start with Theorem 3.1. We will prove the theorem in steps. We will first show that, under the assumptions we have, a unique solution exists on the interval $[0, \delta]$. Then we will show that this solution also satisfies the assumptions we had regarding the initial conditions on the interval $[-\delta, 0]$. This allows us to repeat our arguments to ind a solution on the interval $[\delta, 2\delta]$. Since $\delta > 0$ is fixed, we can iterate this argument until we find a solution on the whole interval $[0, T]$, proving the theorem.

Proof (Theorem 3.1). Consider the process $X(t) = (X^1(t), X^2(t), \ldots, X^N(t))^T$. Then $X(t)$ satisfies the SDE

$$\mathrm{d}X(t) = F(X(t))\mathrm{d}t + G(X(t-\delta))\mathrm{d}t + \varepsilon \mathrm{d}B(t)$$

with

$$F(X(t)) = \begin{pmatrix} f(X^1(t)) \\ f(X^2(t)) \\ \vdots \\ f(X^N(t)) \end{pmatrix}, \quad G(X(t-\delta)) = \begin{pmatrix} g\left(\frac{1}{N}\sum_{j=1}^N X^j(t-\delta)\right) \\ g\left(\frac{1}{N}\sum_{j=1}^N X^j(t-\delta)\right) \\ \vdots \\ g\left(\frac{1}{N}\sum_{j=1}^N X^j(t-\delta)\right) \end{pmatrix}, \quad B(t) = \begin{pmatrix} B^1(t) \\ B^2(t) \\ \vdots \\ B^N(t) \end{pmatrix}.$$

Since the interaction happens with a delay δ, on the interval $[0, \delta]$ we have that $G(X(t-\delta)) = H(t)$ for some functions $H \in C([0, \delta], \mathbb{R}^{dN})$. Since the function f is locally Lipschitz, we also have that the function F is locally Lipschitz, and therefore also the function $F + H$. We also have that B is a dN-dimensional Brownian motion, as each of the Brownian motions B^i are independent. Since we assumed that the initial conditions are $\mathcal{F}_0$-measurable we have that the function $F + H$ is also adapted.

Therefore, by Theorem 2.20, we have a unique solution $\{X(t)\}$ up to time T_∞, where $T_\infty = \lim_{R\to\infty} T_R$. Here T_R is the first time the process X leaves the ball $B(0, R)$ (in dimension dN). We now only need to show that $T_\infty \wedge \delta = \delta$ a.s. We will be showing this component-wise. In particular, let τ_R^i be the first time the process X^i leaves the ball $B(0, R)$ (in dimension d). We will show for each i that $\tau_\infty^i \wedge \delta = \delta$ a.s., with $\tau_\infty^i = \lim_{R\to\infty} \tau_R^i$. Since the amount of particles N is fixed, this clearly implies that $T_\infty \wedge \delta = \delta$ a.s. as well. To show that $\tau_\infty^i \wedge \delta = \delta$, let $1 \le i \le N$ and let $R > 0$. Consider the system of SDEs

$$\begin{aligned} &\mathrm{d}X_R^i(t) = f_R(X_R^i(t))\mathrm{d}t + g\left(\frac{1}{N}\sum_{j=1}^N X_R^j(t-\delta)\right)\mathrm{d}t + \varepsilon \mathrm{d}B^i(t), \quad i \in 1, \ldots, N, t \in [0, \delta], \\ &X_R^i(0) = x_0^i, \\ &X_R^i(s) = \xi_s^i, \quad s \in [-\delta, s), \end{aligned}$$

where f_R is such that $f_R = f$ on $B(0, R)$, is globally Lipschitz and satisfies condition 1 and 2 of Assumption 1. Note that the function g is already globally Lipschitz, so we do not need to make a local approximation. By Theorem 2.19 we have that the exit times of $B(0, R)$ coincide for X^i and X_R^i. Using Itô's formula we compute

$$\begin{aligned} |X_R^i(t)|^2 = |x_0^i|^2 &+ \int_0^t 2\langle X_R^i(s), f_R(X_R^i(s))\rangle + 2\langle X_R^i(s), g\left(\frac{1}{N}\sum_{j=1}^N X_R^j(s-\delta)\right)\rangle \,\mathrm{d}s \\ &+ \int_0^t d\varepsilon^2 \,\mathrm{d}s + \int_0^t 2\varepsilon X_R^i(s)\,\mathrm{d}B(s). \end{aligned}$$

Applying the one-sided Lipschitz to the first inner-product, and Cauchy-Schwarz, the global Lipschitz assumption and Young's inequality to the second inner-product gives

$$\leq |x_0^i|^2 + \int_0^t 2C|X_R^i(s)|^2 + L|X_R^i(s)|^2 + L|\frac{1}{N}\sum_{j=1}^N X_R^j(s-\delta)|^2 + d\varepsilon^2 \,\mathrm{d}s$$
$$+ \int_0^t 2\varepsilon X_R^i(s)\,\mathrm{d}B(s).$$

Taking the supremum over time and expectation yields

$$\mathbb{E}\left[\sup_{0\leq u\leq t}|X_R^i(u)|^2\right] \leq \mathbb{E}\left[|x_0^i|^2\right] + \int_0^t (2C+L)\mathbb{E}\left[\sup_{0\leq u\leq s}|X_R^i(u)|^2\right] + L\frac{1}{N}\sum_{j=1}^n \mathbb{E}\left[|X_R^j(s-\delta)|^2\right]\,\mathrm{d}s$$
$$+ td\varepsilon^2 + \mathbb{E}\left[\sup_{0\leq u\leq t}\int_0^u 2\varepsilon X_R^i(s)\,\mathrm{d}B(s)\right].$$

Using the Burkholder-Davis-Gundy inequality gives

$$\leq \mathbb{E}\left[|x_0^i|^2\right] + \int_0^t (2C+L)\mathbb{E}\left[\sup_{0\leq u\leq s}|X_R^i(u)|^2\right] + L\frac{1}{N}\sum_{j=1}^n \mathbb{E}\left[|X_R^j(s-\delta)|^2\right]\,\mathrm{d}s$$
$$+ td\varepsilon^2 + \mathbb{E}\left[|\int_0^t 4\varepsilon^2|X_R^i(s)|^2\,\mathrm{d}s|^{\frac{1}{2}}\right].$$

Using symmetry of the equations such that all X^j are equally distributed and the fact that $\sqrt{x} \leq \frac{1}{2}(1+x)$ for all $x \in \mathbb{R}_+$ gives

$$\leq \mathbb{E}\left[|x_0^i|^2\right] + T(L\sup_{-\delta\leq u\leq 0}\mathbb{E}\left[|X^i(u)|^2\right] + d\varepsilon^2) + \frac{1}{2}$$
$$+ \int_0^t (2C+L+2\varepsilon^2)\mathbb{E}\left[\sup_{0\leq u\leq s}|X_R^i(u)|^2\right]\,\mathrm{d}s.$$

Using Grönwall's lemma now yields

$$\mathbb{E}\left[\sup_{0\leq t\leq\delta}|X_R^i(t)|^2\right] \leq \left(\mathbb{E}\left[|x_0^i|^2\right] + T(L\sup_{-\delta\leq u\leq 0}\mathbb{E}\left[|X^i(u)|^2\right] + d\varepsilon^2) + \frac{1}{2}\right)e^{(2C+L+2\varepsilon^2)T}.$$

We have that

$$\mathbb{P}\left(\tau_R^i < \delta\right)R^2 \leq \mathbb{E}\left[\sup_{0\leq t\leq\delta}|X_R^i(t)|^2\right].$$

Since we just showed that the right hand side is bounded independent of R, sending R to infinity gives $\mathbb{P}\left(\tau_R^i < \delta\right) \to 0$. This means that $\tau_R^i \to \delta$ a.s., proving existence on the whole interval $[0,\delta]$. Furthermore, using Fatou's lemma the above bound also holds for the process X^i itself. Therefore we have that

$$\sup_{0\leq t\leq\delta}\mathbb{E}\left[|X^i(t)|^2\right] \leq \mathbb{E}\left[\sup_{0\leq t\leq\delta}|X^i(t)|^2\right] < \infty,$$

meaning that the solution on the interval $[0,\delta]$ can serve as an initial condition for a solution on the interval $[\delta, 2\delta]$. Repeating this argument shows existence and uniqueness on the whole interval $[0,T]$, proving the theorem. ■

We continue with the proof of Theorem 3.2. In the proof of Theorem 3.1 we have already seen that the result holds for $p = 2$, even when g is not bounded. We will now present the case when $p > 2$. The proof is inspired by the proof of Theorem 2.3.3 of Reiß (2007).

Proof (Theorem 3.2). Let $1 \leq N$ and $p > 2$. We wish to show that $\mathbb{E}\left[\sup_{0\leq t\leq T}|X^i(t)|^p\right] < \infty$. We will be doing this by showing that there exists a K such that $e^{-K(t\wedge\tau_R)}(1 + |X^i(t\wedge\tau_R)|^p)$ is a supermartingale. Here τ_R is the first time the process X^i leaves the ball $B(0, R)$. We can then use this result to obtain an upper-bound on $\mathbb{P}(\tau_R \leq T)$. With this bound we can derive the required result. To be precise, let

$$K = pC + \frac{p}{2}\left(1 + M^2 + (p - 2 + d)\varepsilon^2\right).$$

By Itô's lemma we have

$$\begin{aligned}
&e^{-Kt}(1 + |X^i(t)|^p) - (1 + |X^i(0)|^p) \\
&= -K\int_0^t e^{-Ks}(1 + |X^i(s)|^p)\,\mathrm{d}s + \int_0^t e^{-Ks}p|X^i(s)|^{p-2}X^i(s)^T\,\mathrm{d}X^i(s) \\
&\quad + \frac{1}{2}\int_0^t e^{-Ks}p(p-2+d)\varepsilon^2|X^i(s)|^{p-2}\,\mathrm{d}s.
\end{aligned}$$

Working out the term $\mathrm{d}X^i(s)$ gives

$$\begin{aligned}
&= \text{ local martingale } + \int_0^t e^{-Ks}\left(-K(1 + |X^i(s)|^p) + p|X^i(s)|^{p-2}\langle X^i(s), f(X^i(s))\rangle\right)\mathrm{d}s \\
&\quad + \int_0^t e^{-Ks}\left(p|X^i(s)|^{p-2}\langle X^i(s), g\left(\frac{1}{N}\sum_{j=1}^N X^j(s-\delta)\right)\rangle + \frac{p}{2}(p-2+d)\varepsilon^2|X^i(s)|^{p-2}\right)\mathrm{d}s.
\end{aligned}$$

Applying the one sided Lipschitz assumption of f to the first inner-product and Cauchy-Schwarz, the boundedness assumption of g and Young's inequality to the second inner product yields

$$\begin{aligned}
&\leq \text{ local martingale } + \int_0^t e^{-Ks}\left(-K(1 + |X^i|^p) + pC|X^i(s)|^p + \frac{p}{2}|X^i(s)|^p\right)\mathrm{d}s \\
&\quad + \int_0^t e^{-Ks}\left(\frac{p}{2}M^2 + \frac{p}{2}(p-2+d)\varepsilon^2|X^i(s)|^{p-2}\right)\mathrm{d}s.
\end{aligned}$$

Using the fact that $|x|^{p-2} \leq 1 + |x|^p$ for all $x \in \mathbb{R}^d$ and grouping everything together gives

$$\leq \text{ local martingale } + \int_0^t e^{-Ks}\left(-K + pC + \frac{p}{2} + \frac{p}{2}M^2 + \frac{p}{2}(p-2+d)\varepsilon^2\right)\left(1 + |X^i(s)|^p\right)\mathrm{d}s.$$

Due to our choice of K we find that $\{e^{-K(t\wedge\tau_R)}(1+|X^i(t\wedge\tau_R)|^p)\}$ is a supermartingale. Therefore we have that

$$\mathbb{E}\left[1 + |X^i(0)|^p\right] \geq \mathbb{E}\left[e^{-K(t\wedge\tau_R)}(1 + |X^i(t\wedge\tau_R)|^p)\right]$$

In particular, this implies that

$$\mathbb{E}\left[1 + |X^i(0)|^p\right]e^{Kt}\frac{1}{1 + R^p} \geq \mathbb{P}(\tau_R \leq t).$$

We now only need to show that this result implies that the moments of the process X^i in the supremum norm are finite. To do this, let $q < p$. We obtain

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le t\le T}|X(s)|^q\right] &= \int_0^\infty \mathbb{P}\left(\sup_{0\le t\le T}|X(s)|^q \ge R\right)\mathrm{d}R\\ &= \int_0^\infty \mathbb{P}\left(\tau_{\sqrt[q]{R}} \le T\right)\mathrm{d}R.\end{aligned}$$

Using the previous result

$$\begin{aligned}&\le \mathbb{E}\left[1+|X^i(0)|^p\right]e^{KT}\left(1+\int_1^\infty \frac{1}{R^{p/q}}\,\mathrm{d}R\right)\\ &= \mathbb{E}\left[1+|X^i(0)|^p\right]e^{KT}\left(1+\frac{q}{p-q}\right) < \infty,\end{aligned}$$

proving the theorem. ■

We now aim to prove Theorem 3.3 and Theorem 3.4, regarding the existence and uniqueness of solutions to the Mckean-Vlasov equation, as well as its moments in the supremum norm. We start out with the proof of Theorem 3.3. The idea is similar to the proof showing that the interacting particle system has unique solutions. We again first show that local existence holds, and then also verify that these solutions in fact exist on the entire interval $[0, T]$. Just as in the proof of Theorem 3.1, we will be doing this step-wise, by first showing that a unique solution exists on the interval $[0, \delta]$, and then showing that this solution satisfies the assumptions we placed on the initial conditions, such that we can iterate this approach to obtain solutions on the intervals $[\delta, 2\delta]$, $[2\delta, 3\delta]$ until we have a solution on the entire interval $[0, T]$. To be precise, the proof is as follows.

Proof (Theorem 3.3). Consider the SDE (3.5) on the interval $[0, \delta]$. Setting $b(x, t) = f(x) + g(\xi_{t-\delta})$, due to Assumption 1, we have that Theorem 2.20 applies, such that we have a solution on the interval $[0, \tau_\infty]$, where $\tau_\infty = \lim_{R\to\infty}\tau_R$, with τ_R the exit time of the ball $B(0, R)$, $R > 0$. We now wish to show that $\lim_{R\to\infty}\tau_R \wedge \delta = \delta$ a.s. Consider the SDE

$$\begin{aligned}\mathrm{d}X_R(t) &= f_R(X_R(t))\mathrm{d}t + g\left(\xi_{t-\delta}\right)\mathrm{d}t + \varepsilon\mathrm{d}B(t), t\in[0,\delta],\\ X_R(0) &= x_0,\\ X_R(s) &= \xi_s, \quad s\in[-\delta, s),\end{aligned}$$

where f_R is such that $f_R = f$ on $B(0, R)$, is globally Lipschitz and satisfies condition 1 and 2 of Assumption 1. Note that the function g is already globally Lipschitz, so we do not need to make a local approximation. By Theorem 2.19 we have that the exit times of the ball $B(0, R)$ coincide for the processes X and X_R.

Using Itô's formula we compute

$$\begin{aligned}|X_R(t)|^2 = |x_0|^2 &+ \int_0^t 2\langle X_R(s), f_R(X_R(s))\rangle + 2\langle X_R(s), g\left(\xi_{s-\delta}\right)\rangle\rangle + d\varepsilon^2\,\mathrm{d}s\\ &+ \int_0^t 2\varepsilon X_R(s)\,\mathrm{d}B(s).\end{aligned}$$

Applying the one-sided Lipschitz to the first inner-product, and Cauchy-Schwarz, the global Lipschitz assumption and Young's inequality to the second inner-product gives

$$\begin{aligned}\le |x_0|^2 &+ \int_0^t 2C|X_R(s)|^2 + L|X_R(s)|^2 + L|\xi_{s-\delta}|^2 + d\varepsilon^2\,\mathrm{d}s\\ &+ \int_0^t 2\varepsilon X_R(s)\,\mathrm{d}B(s).\end{aligned}$$

Taking the supremum over time and expectation yields

$$\mathbb{E}\left[\sup_{0\le u\le t}|X_R(u)|^2\right] \le \mathbb{E}\left[|x_0|^2\right] + \int_0^t (2C+L)\mathbb{E}\left[\sup_{0\le u\le s}|X_R(u)|^2\right] + L|\xi_{s-\delta}|^2\,\mathrm{d}s$$
$$+\, td\varepsilon^2 + \mathbb{E}\left[\sup_{0\le u\le t}\int_0^u 2\varepsilon X_R(s)\,\mathrm{d}B(s)\right].$$

Using the Burkholder-Davis-Gundy inequality

$$\le \mathbb{E}\left[|x_0|^2\right] + \int_0^t (2C+L)\mathbb{E}\left[\sup_{0\le u\le s}|X_R(u)|^2\right] + L|\xi_{s-\delta}|^2\,\mathrm{d}s$$
$$+\, td\varepsilon^2 + \mathbb{E}\left[|\int_0^t 4\varepsilon^2|X_R(s)|^2\,\mathrm{d}s|^{\frac{1}{2}}\right].$$

Using the fact that $\sqrt{x} \le \frac{1}{2}(1+x)$ for all $x \in \mathbb{R}_+$ gives

$$\le \mathbb{E}\left[|x_0|^2\right] + T(L\sup_{-\delta\le u\le 0}|\xi_u| + d\varepsilon^2) + \frac{1}{2}$$
$$+ \int_0^t (2C+L+2\varepsilon^2)\mathbb{E}\left[\sup_{0\le u\le s}|X_R(u)|^2\right]\,\mathrm{d}s.$$

Using Grönwall's lemma now yields

$$\mathbb{E}\left[\sup_{0\le t\le \delta}|X_R(t)|^2\right] \le \left(\mathbb{E}\left[|x_0|^2\right] + T(L\sup_{-\delta\le u\le 0}|\xi_u| + d\varepsilon^2) + \frac{1}{2}\right)e^{(2C+L+2\varepsilon^2)T}.$$

We have that

$$\mathbb{P}\left(\tau_R < \delta\right)R^2 \le \mathbb{E}\left[\sup_{0\le t\le\delta}|X_R(t)|^2\right].$$

We now continue in the same way as we did in the proof of Theorem 3.1. In particular, we argue as follows. Since we just showed that the right hand side is bounded independent of R, sending R to infinity gives $\mathbb{P}\left(\tau_R < \delta\right) \to 0$. This means that $\tau_R \to \delta$ a.s., proving existence on the whole interval $[0,\delta]$. Furthermore, using Fatou's lemma the above bound also holds for the process X itself. Therefore we have that

$$\sup_{0\le t\le\delta}\mathbb{E}\left[|X(t)|^2\right] \le \mathbb{E}\left[\sup_{0\le t\le\delta}|X(t)|^2\right] < \infty,$$

meaning that the solution on the interval $[0,\delta]$ can serve as an initial condition for a solution on the interval $[\delta,2\delta]$. Repeating this argument shows existence and uniqueness on the whole interval $[0,T]$, proving the theorem. ■

Lastly, we present the proof of Theorem 3.4. We already have proven the result in the case that $p \le 2$. For the proof of the case $p \ge 2$ we use a similar approach as we did for the proof of Theorem 3.2. That is, we show that there exists a K such that ${}^{-Kt}(1+|X_R(t)|^p)$ is a super-martingale. This will give us an upper-bound on the probability that $\tau_R < T$, in a way that will allow us to show that the moments of X in the supremum norm are finite. This time, however, we do not need the assumption that g is bounded. This is because we have a deterministic upper-bound on the expectation of X at time $t, t \in [0,T]$, while for the interacting case we did not have a deterministic upper-bound for the empirical mean of the particles. The precise proof is as follows.

Proof (Theorem 3.4). Let $p \geq 2$. Let

$$K = pC + \frac{p}{2}L + \frac{p}{2}(p-2+d)\varepsilon^2 \\ + \frac{p}{2}L\left(\sup_{-\delta\leq r\leq 0}\mathbb{E}\left[|X(r)|^2\right] + \left(\mathbb{E}\left[|X(0)|^2\right] + TL\sup_{-\delta\leq r\leq 0}\mathbb{E}\left[|X(r)|^2\right] + T\varepsilon^2 + \frac{1}{2}\right)e^{(2C+L)T}\right).$$

Using Itô's lemma we compute

$$e^{-Kt}(1+|X(t)|^p) - (1+|X(0)|^p) \\ = -K\int_0^t e^{-Ks}(1+|X(s)|^p)\,\mathrm{d}s + \int_0^t e^{-Ks}p|X(s)|^{p-2}X(s)^T\,\mathrm{d}X(s) \\ + \frac{1}{2}\int_0^t e^{-Ks}p(p-2+d)\varepsilon^2|X(s)|^{p-2}\,\mathrm{d}s.$$

Expending the term $\mathrm{d}X(s)$ yields

$$= \text{ local martingale } + \int_0^t e^{-Ks}\left(-K(1+|X(s)|^p) + p|X(s)|^{p-2}\langle X(s), f(X(s))\rangle\right)\,\mathrm{d}s \\ + \int_0^t e^{-Ks}\left(p|X(s)|^{p-2}\langle X(s), g(\mathbb{E}\left[X(s-\delta)\right])\rangle + \frac{p}{2}(p-2+d)\varepsilon^2|X(s)|^{p-2}\right)\,\mathrm{d}s.$$

Applying the one-sided Lipschitz assumption on f to the first inner-product and Cauchy-Schwarz, the global Lipschitz assumption on g and Young's inequality gives us

$$\leq \text{ local martingale } + \int_0^t e^{-Ks}\left(-K(1+|X|^p) + pC|X(s)|^p + \frac{p}{2}L|X(s)|^2\right)\,\mathrm{d}s \\ + \int_0^t e^{-Ks}\left(\frac{p}{2}L\mathbb{E}\left[|X(s-\delta)|^2\right] + \frac{p}{2}(p-2+d)\varepsilon^2|X(s)|^{p-2}\right)\,\mathrm{d}s.$$

Using the fact that $|x|^{p-2} \leq 1 + |x|^p$ for all $x \in \mathbb{R}^d$ and reordering all the terms yields

$$\leq \text{ local martingale} \\ + \int_0^t e^{-Ks}\left(-K + pC + \frac{p}{2}L + \frac{p}{2}L\mathbb{E}\left[|X(s-\delta)|^2\right] + \frac{p}{2}(p-2+d)\varepsilon^2\right)(1+|X(s)|^p)\,\mathrm{d}s.$$

By our choice of K, and the bound on the second moment we found earlier, we find that $\{e^{-K(t\wedge\tau_R)}(1+|X(t\wedge\tau_R)|^p)\}_{0\leq t\leq T}$ is a supermartingale. Therefore we have that

$$\mathbb{E}\left[1+|X(0)|^p\right] \geq \mathbb{E}\left[e^{-K(t\wedge\tau_R)}(1+|X(t\wedge\tau_R)|^p)\right].$$

In particular we have that

$$\mathbb{E}\left[1+|X(0)|^p\right]e^{Kt}\frac{1}{1+R^p} \geq \mathbb{P}(\tau_R \leq t). \tag{3.9}$$

Lastly, we show how this result implies that the moments of X in the supremum norm are finite. Let $q < p$. We obtain

$$\mathbb{E}\left[\sup_{0\leq s\leq T}|X(s)|^q\right] = \int_0^\infty \mathbb{P}(\sup_{0\leq s\leq T}|X(s)|^q \geq R)\,\mathrm{d}R \\ = \int_0^\infty \mathbb{P}(\tau_{\sqrt[q]{R}} \leq T)\,\mathrm{d}R.$$

Using the previous result gives

$$
\begin{aligned}
&\leq \mathbb{E}\left[1 + |X(0)|^p\right] e^{KT} \left(1 + \int_1^\infty \frac{1}{R^{p/q}} \,\mathrm{d}R\right) \\
&= \mathbb{E}\left[1 + |X(0)|^p\right] e^{KT} \left(1 + \frac{q}{p-q}\right) < \infty,
\end{aligned}
$$

proving the theorem. ■

Remark. *Instead of assuming that g is Lipschitz, we can also assume that g is bounded by some constant $M > 0$, i.e. for all $x \in \mathbb{R}^d$ we have that $|g(x)| < M$. In this case we can improve the value of K to*

$$
K = pC + \frac{p}{2} + \frac{p}{2}(p - 2 + d)\varepsilon^2 + \frac{p}{2}M^2.
$$

The proof for this is similar to the proof of Theorem 3.2, as the assumption of g being bounded was also made there.

4. Propagation of chaos

So far we have been discussing two processes. The first process describes an interacting particle system, where evolution of the N particles depends on the mean position of all the particles with a delay $\delta > 0$. The other processes is a delayed McKean-Vlasov process, where this stochastic empirical mean term is replaced by the (deterministic) expectation of the process itself. By the law of large numbers we know that the empirical mean converges towards the true mean. Therefore it is natural to assume that there is some relation between the two models, and that the McKean-Vlasov process can be seen as the limiting process when the amount of particles N goes to infinity in some manner. In this chapter we will describe precisely in what way this limit holds.

We will find out that, if we create n copies of the McKean-Vlasov process and couple them to the first n particles of the interacting particle process by the means of using the same Brownian motion and initial condition, under certain assumptions, we can show that the distance in supremum norm between the interacting particle process and the McKean-Vlasov process converges to 0 when sending N to infinity, while keeping n fixed. Furthermore, we also obtain an upper bound on the rate of convergence. This result implies that the interacting particle system possesses the so called *propagation of chaos* property. This means that the marginals of the first n particles of the interacting particle process at a time t will converge to the marginals of the n McKean-Vlasov processes, which are all independent of each other given that the initial distributions of these processes are independent. Therefore we have that the chaos (independence) of the initial condition propagates forward to all the marginals in the limit of the amount of particles N going to infinity.

4.1. An easier example

Before we analyse the main model as presented in Section 1.1, we first consider a simpler model to get an idea of the concept. Here we follow Chapter 1 of Sznitman (1991). In particular, we consider, for $N > 0$, the system of SDEs

$$\mathrm{d}X^{i,N}(t) = \frac{1}{N}\sum_{j=1}^{N} b(X^{i,N}(t), X^{j,N}(t))\mathrm{d}t + \mathrm{d}B^i(t), \quad i = 1, \dots, N, \tag{4.1}$$

$$X^i(0) = x_0^i, \quad i = 1, \dots, N. \tag{4.2}$$

Notice that there are three major differences between this model and our model. Firstly, the drift function b is not separated into two functions, depending on $X^{i,N}$ and $X^{j,N}$ respectively, but as one function, allowing for more direct interaction. Secondly, in this model there is no delay in the interaction. Lastly, in this model we consider the average of the interactions, rather than an interaction withthe average. We will see that this also results in a different limiting model. In particular, consider the McKean-Vlasov equation

$$\mathrm{d}\bar{X}^i(t) = x_0^i + \int_0^t \int b(\bar{X}^i(s), y)u_s(\mathrm{d}y)\,\mathrm{d}s + B^i(t), \quad i = 1, \dots, N, \tag{4.3}$$

$$u_s(\mathrm{d}y) = \mathrm{law}(\bar{X}^i(s)), \tag{4.4}$$

where the initial conditions x_0^i and the Brownian motions B^i are the same as those for the interacting particle system. As stated, we now indeed have a different limiting model, as the expectation over the law of $\bar{X}^i(s)$ is now outside the function b rather than inside. The reason that we call this the limiting model is because of the following theorem we have.

Theorem 4.1 (Theorem 1.4 of Sznitman (1991)). *Suppose that b is bounded and Lipschitz in both arguments. Suppose that the initial conditions $x_0^i, i = 1, \dots, N$ are i.i.d. Then for any $i \geq 1, T > 0$ we have*

$$\sup_{N \geq i} \sqrt{N} \mathbb{E} \left[\sup_{0 \leq t \leq T} |X^{i,N}(t) - \bar{X}^i(t)| \right] < \infty. \tag{4.5}$$

We omit the proof here, as it is similar to the proof of Theorem 4.3, which we will present later in this chapter.

This theorem tells us that when N grows large, the expected distance in the supremum norm between the trajectories of the interacting particle model and the trajectories of the (independent) McKean-Vlasov process becomes small.

We now wish to show the propagation of chaos property. We first remark that by Proposition 2.2 of Sznitman (1991) we only have to consider two particles instead of n particles, and show that the propagation of chaos result holds for this pair. This pairwise independence implies non-trivially that the propagation of chaos also holds for a set of n particles, with $n \geq 2$ fixed.

Using Theorem 4.1, we can easily show that $X^{i,N} \to \bar{X}^i$ in probability. Indeed, let $\theta > 0$. By Markov's inequality and Theorem 4.1 we have that there exists a constant $M > 0$ such that

$$\mathbb{P} \left(\sup_{0 \leq t \leq T} |X^{i,N}(t) - \bar{X}^i(t)| > \theta \right) \leq \frac{1}{\theta} \mathbb{E} \left[\sup_{0 \leq t \leq T} |X^{i,N}(t) - \bar{X}^i(t)| \right] \leq \frac{M}{\theta} \frac{1}{\sqrt{N}}. \tag{4.6}$$

Sending $N \to \infty$ we see that we indeed have convergence in probability in the space $C([0,T], \mathbb{R}^d)$ equipped with the supremum norm.

We now define what we mean when we say that two sequences of random variables are asymptotically independent.

Definition 4.2. *Let $(X_n, Y_n)_n$ be a sequence of pairs of random variables, converging in distribution to the pair (X, Y). We call the sequence (X_n, Y_n) asymptotically independent if the pair (X, Y) is independent.*

To show that we have propagation of chaos, which is equivalent to saying that for $i \neq j$ we have that the pair $(X^{i,N}, X^{j,N})_N$ is asymptotically independent, we want to show that the pair $(X^{i,N}, X^{j,N})_N$ converges in distribution to the pair $(\bar{X}^i, \bar{X}^j)$ when $N \to \infty$, which is independent by construction. We will show that this convergence even holds in probability. Indeed, let $\theta > 0$, we have

$$\begin{aligned}
&\mathbb{P} \left(\|(X^{i,N}, Y^{i,N}) - (\bar{X}^i, \bar{X}^j)\| > \theta \right) \\
&\leq \mathbb{P} \left(\sup_{0 \leq t \leq T} |X^{i,N}(t) - \bar{X}^i(t)| + \sup_{0 \leq t \leq T} |X^{j,N}(t) - \bar{X}^j(t)| > \theta \right) \\
&\leq \mathbb{P} \left(\sup_{0 \leq t \leq T} |X^{i,N}(t) - \bar{X}^i(t)| > \frac{\theta}{2} \cup \sup_{0 \leq t \leq T} |X^{j,N}(t) - \bar{X}^j(t)| > \frac{\theta}{2} \right) \\
&\leq \mathbb{P} \left(\sup_{0 \leq t \leq T} |X^{i,N}(t) - \bar{X}^i(t)| > \frac{\theta}{2} \right) + \mathbb{P} \left(\sup_{0 \leq t \leq T} |X^{j,N}(t) - \bar{X}^j(t)| > \frac{\theta}{2} \right),
\end{aligned}$$

where the norm in the first line denotes the supremum norm on the product space $C([0,T], \mathbb{R}^d)^2$. The terms in the last line go to 0 by Equation (4.6). Thus we have that $(X^{i,N}, X^{j,N})_N$ converges in probability to the pair $(\bar{X}^i, \bar{X}^j)$ when $N \to \infty$. Since convergence in probability implies convergence in distribution, we have that the pair $(X^{i,N}, X^{j,N})_N$ is asymptotically independent, as the pair $(\bar{X}^i, \bar{X}^j)$ is independent.

4.2. Interaction with a delay

We now wish to apply similar arguments as above to the model we introduced in Section 1.1. Namely, we consider the following model. For $N \geq 2$ and $\delta > 0$, consider the interacting particle process

$$\begin{aligned} \mathrm{d}X^{i,N}(t) &= f(X^{i,N}(t))\mathrm{d}t + g\left(\frac{1}{N}\sum_{j=1}^{N} X^{j,N}(t-\delta)\right)\mathrm{d}t + \varepsilon \mathrm{d}B^i(t), \quad i = 1, \dots, N, \\ X^{i,N}(0) &= x_0^i, \\ X^{i,N}(s) &= \xi(s), \quad s \in [-\delta, 0), \end{aligned}$$

with B^i independent standard Brownian motions. Also consider the processes

$$\begin{aligned} \mathrm{d}\bar{X}^i(t) &= f(\bar{X}^i(t))\mathrm{d}t + g\left(\mathbb{E}\left[\bar{X}^i(t-\delta)\right]\right)\mathrm{d}t + \varepsilon \mathrm{d}B^i(t), \quad i = 1, \dots, N, \\ \bar{X}^i(0) &= x_0^i, \\ X^i(s) &= \xi(s), \quad s \in [-\delta, 0), \end{aligned}$$

where the initial conditions and the Brownian motions are the same as for the interacting particle system. As stated before, the model differs from the previously discussed model in several ways. Firstly the structure of the interaction is different. Instead of considering one function depending on both the current particle and the particle with which it interacts, we now have two functions, f and g, which consider the drift of the particle itself and its interaction with the other particles separately.

Secondly, the interaction now happens as a function of the mean, rather than the mean of a function. This is also reflected in the McKean-Vlasov equation, where the expectation is inside the function g, rather than outside.

Lastly, the interaction now occurs with a delay δ. Although this changes the model significantly, we will see that the results concerning the propagation of chaos do not change because of this. The reason for this is that we have a bound on the second moment of the process uniformly in time. To be precise, we present the following theorem, which can be seen as the equivalent of Theorem 4.1.

Theorem 4.3. *Assume that Assumption 1 holds. Assume that the initial conditions $\xi(s), s \in [-\delta, 0)$ and $x_0^i, i = 1, \dots, N$ are i.i.d. We then have*

$$\sup_{N \geq i} N\mathbb{E}\left[\sup_{0 \leq t \leq T} |X^{i,N}(t) - \bar{X}^i(t)|^2\right] < \infty.$$

If we compare Theorem 4.3 to Theorem 4.1 we can note some differences. Firstly, the result is slightly stronger, as this time we consider the second moment rather than the first moment. Since $\mathbb{E}[X]^2 \leq \mathbb{E}[X^2]$, we can obtain results for the first moment based on the second moment, but not the other way around.

Also the required assumptions for Theorem 4.3 are milder than those for Theorem 4.1. The latter has a global Lipschitz assumption on both arguments. Furthermore, it assumes that the function b is bounded. In this theorem the assumptions are weakened to require only a global Lipschitz property for the function g, and only a local Lipschitz with a growth condition for the function f. Also neither of these functions are assumed to be bounded.

Before we present the proof of the theorem, we first present a small algebraic lemma, which we will use in the proof.

Lemma 4.4.

$$a\frac{1}{N}\sum_{j=1}^{n} b_j \le \frac{1}{2}a^2 + \frac{1}{2N}\sum_{j=1}^{N} b_j^2.$$

Proof.

$$\begin{aligned} a\frac{1}{N}\sum_{j=1}^{N} b_j &= \sum_{j=1}^{N} \frac{a}{\sqrt{N}}\frac{b_j}{\sqrt{N}} \\ &\le \sum_{j=1}^{N} \frac{a^2}{2N} + \frac{b_j^2}{2N} = \frac{1}{2}a^2 + \frac{1}{2N}\sum_{j=1}^{N} b_j^2. \end{aligned}$$

■

We now present the proof of theorem. We will show that the squared distance between X^i and $\bar{X}^i$ can be related to the variance of the empirical mean of the processes $\bar{X}^i, i = 1, \dots, N$. Since these processes are independent, we have that this variance decays linearly in N, which will allow us to prove the result.

Proof. Using Itô's formula we compute

$$\begin{aligned} \mathrm{d}|X^i(t) - \bar{X}^i(t)|^2 = {} & 2\left\langle X^i(t) - \bar{X}^i(t), f(X^i(t)) - f(\bar{X}^i(t))\right\rangle \mathrm{d}t \\ & + 2\left\langle X^i(t) - \bar{X}^i(t), g\left(\frac{1}{N}\sum_{j=1}^{N} X^j(t-\delta)\right) - g\left(\mathbb{E}\left[\bar{X}^i(t-\delta)\right]\right)\right\rangle \mathrm{d}t. \end{aligned}$$

Or equivalently

$$\begin{aligned} |X^i(t) - \bar{X}^i(t)|^2 = {} & \int_0^t 2\left\langle X^i(s) - \bar{X}^i(s), f(X^i(s)) - f(\bar{X}^i(s))\right\rangle \mathrm{d}s \\ & + \int_0^t \left\langle X^i(s) - \bar{X}^i(s), g\left(\frac{1}{N}\sum_{i=j}^{N} X^j(s-\delta)\right) - g\left(\mathbb{E}\left[\bar{X}^i(s-\delta)\right]\right)\right\rangle \mathrm{d}s \end{aligned}$$

Applying the one-sided Lipschitz assumption of f to the first inner-product, and Cauchy-Schwarz and the global Lipschitz assumption of g to the second inner-product we obtain

$$\begin{aligned} \le {} & \int_0^t 2C|X^i(s) - \bar{X}^i(s)|^2 \,\mathrm{d}s \\ & + \int_0^t |X^i(s) - \bar{X}^i(s)| L \left|\frac{1}{N}\sum_{j=1}^{N} X^j(s-\delta) - \mathbb{E}\left[\bar{X}^i(s-\delta)\right]\right| \mathrm{d}s \\ \le {} & \int_0^t 2C|X^i(s) - \bar{X}^i(s)|^2 \,\mathrm{d}s \\ & + \int_0^t L|X^i(s) - \bar{X}^i(s)| \left|\frac{1}{N}\sum_{j=1}^{N} X^j(s-\delta) - \bar{X}^j(s-\delta)\right| \mathrm{d}s \\ & \quad + \int_0^t L|X^i(s) - \bar{X}^i(s)| \left|\frac{1}{N}\sum_{j=1}^{N} \bar{X}^j(s-\delta) - \mathbb{E}\left[\bar{X}^j(s-\delta)\right]\right| \mathrm{d}s. \end{aligned}$$

Using Lemma 4.4

$$
\begin{aligned}
&\leq (2C + L^2) \int_0^t |X^i(s) - \bar{X}^i(s)|^2 \, \mathrm{d}s \\
&\quad + \int_0^t \frac{1}{2N} \sum_{j=1}^N |X^j(s-\delta) - \bar{X}^j(s-\delta)|^2 \, \mathrm{d}s \\
&\quad + \frac{1}{2} \int_0^t \left| \frac{1}{N} \sum_{j=1}^N \bar{X}^j(s-\delta) - \mathbb{E}\left[\bar{X}^j(s-\delta)\right] \right|^2 \mathrm{d}s.
\end{aligned}
$$

Next we take the supremum over time and expectation. Noting that $X^j - \bar{X}^j$ is identically distributed for each j by symmetry, we obtain

$$
\begin{aligned}
\mathbb{E}\left[\sup_{0\leq r\leq t} |X^i(r) - \bar{X}^i(r)|^2\right] &\leq (2C + L^2 + \frac{1}{2}) \int_0^t \mathbb{E}\left[\sup_{0\leq r\leq s} |X^i(r) - \bar{X}^i(r)|^2\right] \mathrm{d}s \\
&\quad + \frac{1}{2} \int_0^t \mathbb{E}\left[\left| \frac{1}{N} \sum_{j=1}^N \bar{X}^j(s-\delta) - \mathbb{E}\left[\bar{X}^j(s-\delta)\right] \right|^2\right] \mathrm{d}s.
\end{aligned}
$$

Applying Grönwall's inequality yields

$$
\mathbb{E}\left[\sup_{0\leq r\leq T} |X^i(r) - \bar{X}^i(r)|^2\right] \leq \frac{1}{2} e^{(2C+L^2+\frac{1}{2})T} \int_0^T \mathbb{E}\left[\left| \frac{1}{N} \sum_{j=1}^N \bar{X}^j(s-\delta) - \mathbb{E}\left[\bar{X}^j(s-\delta)\right] \right|^2\right] \mathrm{d}s.
$$

Notice that we have

$$
\begin{aligned}
&\mathbb{E}\left[\left| \frac{1}{N} \sum_{j=1}^N \bar{X}^j(s-\delta) - \mathbb{E}\left[\bar{X}^j(s-\delta)\right] \right|^2\right] \\
&= \frac{1}{N^2} \mathbb{E}\left[\sum_{l=1}^d \sum_{j,k=1}^N \left(\bar{X}^j_l(s-\delta) - \mathbb{E}\left[\bar{X}^j_l(s-\delta)\right]\right)\left(\bar{X}^k_l(s-\delta) - \mathbb{E}\left[\bar{X}^k_l(s-\delta)\right]\right)\right],
\end{aligned}
$$

where $\bar{X}^j_l(s-\delta)$ denotes the l'th component of $\bar{X}^j$ at time $s-\delta$. Using independence of the $\bar{X}^j$'s we obtain

$$
\begin{aligned}
&= \frac{1}{N^2} \mathbb{E}\left[\sum_{l=1}^d \sum_{j=1}^N \left(\bar{X}^j_l(s-\delta) - \mathbb{E}\left[\bar{X}^j_l(s-\delta)\right]\right)^2\right] \\
&= \frac{1}{N^2} \mathbb{E}\left[\sum_{j=1}^N \left|\bar{X}^j(s-\delta) - \mathbb{E}\left[\bar{X}^j(s-\delta)\right]\right|^2\right].
\end{aligned}
$$

By symmetry of the distributions of $\bar{X}^j$ we have

$$
\begin{aligned}
&= \frac{1}{N} \mathbb{E}\left[\left|\bar{X}^1(s-\delta) - \mathbb{E}\left[\bar{X}^1(s-\delta)\right]\right|^2\right] \\
&\leq \frac{M}{N},
\end{aligned}
$$

for some constant $M > 0$. See Chapter 3 for details on this. Therefore we obtain that

$$
\begin{aligned}
\mathbb{E}\left[\sup_{0\leq t\leq T} |X^i(t) - \bar{X}^i(t)|^2\right] &\leq \frac{1}{2} e^{(2C+L^2+\frac{1}{2})T} \int_0^T \frac{M}{N} \, \mathrm{d}s \\
&= \frac{1}{2} T e^{(2C+L^2+\frac{1}{2})T} \frac{M}{N} = \frac{\tilde{M}}{N},
\end{aligned}
$$

proving the theorem. ■

Now that we have proven Theorem 4.3, we can apply a similar analysis as we did in Section 4.1. We again have that $X^{i,N} \to \bar{X}^i$ in probability in the space $C([0,T], \mathbb{R}^d)$ equipped with the supremum norm. Therefore we also have for any $i \neq j$ that the pair $(X^{i,N}, X^{j,N}) \to (\bar{X}^i, \bar{X}^j)$ in probability as $N \to \infty$. From this we can conclude that the pair is asymptotically independent and that thus the propagation of chaos property also holds for this model, under the assumptions as stated in Theorem 4.3.

5. Small noise regime

In the previous chapter we discussed what happens to our model when we send the amount of particles N to infinity. In particular, we found that for a fixed amount of n particles, the trajectories of interacting particles converged to those of the McKean-Vlasov equation.

In this chapter we will be looking at a different limit. We will now investigate what happens to the model when the noise parameter ε is send to 0. We are especially interested in whether or not a LDP applies to our model. LDPs concern at which rate probabilities of rare events occur on an exponential scale when the noise parameter ε is send to 0. Let us make this more precise. We first define what we call a rate function.

Definition 5.1. *A* rate function *is a lower semi-continuous function $I : X \to [0, \infty]$, i.e. for all $\alpha \in [0, \infty)$, the level sets*

$$\Psi_I(\alpha) = \{x \in X \,:\, I(x) \leq \alpha\}$$

are closed. I is called a good rate function *if all level sets are compact.*

With this definition in hand we can now define what we precisely mean with a LDP.

Definition 5.2. *Let I be a rate function. A family of probability measures $\{\mu_\varepsilon\}_{\varepsilon>0}$ satisfies a* large deviation principle with rate function I *if for all $\Gamma \in \mathcal{B}$ we have*

$$-\inf_{x \in \Gamma^o} I(x) \leq \liminf_{\varepsilon \to 0} \varepsilon \ln \mu_\varepsilon(\Gamma) \leq \limsup_{\varepsilon \to 0} \ln \mu_\varepsilon(\Gamma) \leq -\inf_{x \in \overline{\Gamma}} I(x).$$

Here Γ^o denotes the open interior and $\overline{\Gamma}$ the closure of the set Γ.

Remark. *We often say that a family of random variables $\{X^\varepsilon\}_{\varepsilon>0}$ satisfies a LDP. In this case we mean that the family of laws of the random variables satisfies a LDP.*

As said, in this chapter we investigate if solutions to SDEs satsify a LDP. We will start with Schilder's theorem, describing a LDP for Brownian motions. Using the so called contraction principle, we are then able to show the Freidlin-Wentzell theorem, describing a LDP for SDEs with a Lipschitz drift function and additive noise.

Next we look at a general result by Chiarini and Fischer (2014), describing a LDP for a class of stochastic delay equations, which may not only depend on the current state and the state at a fixed delay δ, but rather on the entire trajectory up until that point.

Finally, we investigate which assumptions we need to make such that we can apply the result of Chiarini and Fischer (2014) to our model.

We will find that under a local Lipschitz assumption with a linear growth condition we are able to show a LDP for both the interacting model as well as for the McKean-Vlasov equation. Without the linear growth condition we are still able to show pathwise convergence of the solution to the corresponding ODE, but we are unable to proof that a LDP holds in that case.

5.1. Freidlin–Wentzell theorem

Throughout this section we will refer to various theorems obtained from Herrmann et al. (2014). We will refer to them for the proofs of the theorems when we do not provide the proofs ourselves. We start by presenting Schilder's theorem, describing a LDP for the Brownian motion re-scaled by the square root of the noise parameter ε. Firstly, we notice that for $t \in [0, T]$ we have that $\sqrt{\varepsilon}B(t)$ is normally distributed with mean 0 and variance εt. We thus expect that the marginals converge to 0 when $\varepsilon \to 0$. The reflection principle of the Brownian motion then implies that this also holds on a trajectory level. A typical trajectory therefore stays close t0 0 trajectory when ε is small. Schilder's theorem now tells us how fast the probability that non-typical trajectories occur, decays when $\varepsilon \to 0$. To be precise, the theorem states the following.

Theorem 5.3 (Schilder's theorem; Theorem 2.28 in Herrmann et al. (2014)). *Let B be a d-dimensional Brownian motion. For $\varepsilon > 0$, let $B^\varepsilon = \sqrt{\varepsilon}B$. Then on the space $C([0, T], \mathbb{R}^d)$ equipped with the supremum norm, the family $\{B^\varepsilon\}_\varepsilon$ satisfies a LDP with good rate function*

$$I(\phi) = \begin{cases} \frac{1}{2}\int_0^T |\dot{\phi}(t)|^2 \,\mathrm{d}t, & \text{if } \phi \in \mathcal{H}_1([0, T], \mathbb{R}^d), \\ +\infty, & \text{else.} \end{cases}$$

Here $\mathcal{H}_1([0, T], \mathbb{R}^d)$ denotes the Cameron-Martin space of absolutely continuous functions, which is defined as

$$\begin{aligned} \mathcal{H}_1([0, T], \mathbb{R}^d) &= \left\{ f : [0, 1] \to \mathbb{R}^d, f(0) = 0, f \text{ is absolutely continuous with } |\dot{f}| \in L^2([0, 1], \mathbb{R}) \right\} \\ &= \left\{ \int_0^t \dot{f}(s) \,\mathrm{d}s, |\dot{f}| \in L^2([0, 1], \mathbb{R}) \right\}. \end{aligned}$$

We will not discuss the proof of this theorem here. It can be found in Section 2.3.3 of Herrmann et al. (2014). We now wish to extend this theorem for Brownian motions to solutions of equations of the following type.

$$\mathrm{d}X^\varepsilon(t) = b(X^\varepsilon(t))\mathrm{d}t + \sqrt{\varepsilon}B(t), \quad t \in [0, T]. \tag{5.1}$$

To do this, we make use of the so called contraction principle. This important theorem in the study of LDPs allows us to derive a LDP for a certain family of random variables based on an existing LDP, given that there exists a continuous mapping from the latter to the former. Precisely, the contraction principle states the following.

Theorem 5.4 (Contraction principle; Theorem 2.17 in Herrmann et al. (2014)). *Let $\mathbf{X}$ and $\mathbf{Y}$ be topological spaces, and $f : \mathbf{X} \to \mathbf{Y}$ a continuous mapping. Let $I : \mathbf{X} \to [0, \infty]$ be a good rate function. The following holds.*

1. *For $y \in \mathbf{Y}$ let*

$$I'(y) = \inf\{I(x) : x \in \mathbf{X}, y = f(x)\}$$

with the convention that $\inf\{\varnothing\} = +\infty$. Then I' is a good rate function on $\mathbf{Y}$;

2. *Suppose that $\{\mu_\varepsilon\}_{\varepsilon>0}$ satisfies a LDP with (good) rate function I, and $\nu_\varepsilon = \mu_\varepsilon \circ f^{-1}$, $\varepsilon > 0$. Then the family $\{\nu_\varepsilon\}_{\varepsilon>0}$ satisfies a LDP with (good) rate function I'.*

We now wish to use this contraction principle to show a LDP for the family of solutions $\{X^\varepsilon\}_{\varepsilon>0}$ which satisfy Equation (5.1). For this we would like to use the space $\mathbf{X}$ as $C([0, T], \mathbb{R}^d)$ equipped with the supremum norm, where the laws of the family $\{\sqrt{\varepsilon}B\}_{\varepsilon>0}$ play the role of the family $\{\mu_\varepsilon\}_{\varepsilon>0}$. For $\mathbf{Y}$ we wish to use the same space as for $\mathbf{X}$, but let $\{\nu_\varepsilon\}_{\varepsilon>0}$ play the role of the laws of the family of solutions $\{X^\varepsilon\}_{\varepsilon>0}$. For this we would need a continuous function f which maps a Brownian motion to a solution of Equation (5.1). This function indeed exists and gives rise to the Freidlin-Wentzell theorem.

Theorem 5.5 (Freidlin-Wentzell theorem; Theorem 2.31 in Herrmann et al. (2014)). *Let X^ε be the solution to the SDE*

$$\begin{aligned} \mathrm{d}X^\varepsilon(t) &= b(X^\varepsilon(t))\mathrm{d}t + \sqrt{\varepsilon}\mathrm{d}B(t), \quad t \in [0,T], \\ X^\varepsilon(0) &= x, \end{aligned}$$

where B denotes a standard $\mathbb{R}^d$-valued Brownian motion. Assume that function $b : \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz with Lipschitz constant L. Then $\{X(t)\}_{\varepsilon>0}$ satisfies a LDP with good rate function

$$I(\phi) = \begin{cases} \frac{1}{2}\int_0^T |\dot{\phi}(t) - b(\phi(t))|^2 \,\mathrm{d}t, & \text{if } \phi - x \in \mathcal{H}_1([0,T],\mathbb{R}^d), \\ +\infty, & \text{else.} \end{cases}$$

Proof. As said before, we wish to show that the mapping

$$F : C([0,1],\mathbb{R}^d) \to C([0,1],\mathbb{R}^d),$$

which maps a function g to a function f which satisfies

$$f(t) = x + \int_0^t b(f(s))\,\mathrm{d}s + g(t), \quad t \in [0,T], \tag{5.2}$$

is well-defined and continuous. We start with the well-definedness. We will show that there exists a unique solution of (5.2) for all $x \in \mathbb{R}^d, g \in C([0,T],\mathbb{R}^d)$ with T such that $LT < 1$. We can then in steps construct a solution of length T for any $T > 0$. To show this existence, define the mapping

$$\begin{aligned} \Gamma : C([0,T],\mathbb{R}^d) &\to C([0,T],\mathbb{R}^d), \\ f(\cdot) &\mapsto x + \int_0^\cdot b(f(s))\,\mathrm{d}s + g(\cdot). \end{aligned}$$

We will show that this mapping is a contraction. Let $f_1, f_2 \in C([0,T],\mathbb{R}^d)$. We have

$$\sup_{0\le t\le T} |\Gamma(f_1)(t) - \Gamma(f_2)(t)| = \int_0^T |b(f_1(s)) - b(f_2(s))|\,\mathrm{d}s.$$

Using the Lipschitz assumption

$$\begin{aligned} &\le \int_0^T L|f_1(s) - f_2(s)|\,\mathrm{d}s \\ &\le \int_0^T L\|f_1 - f_2\|\,\mathrm{d}s \\ &= LT\|f_1 - f_2\|. \end{aligned}$$

Since we assumed that $LT < 1$ we indeed have a contraction. By the Banach fixed point theorem we find a unique solution f. Next we show that F is continuous. Let $g_1, g_2 \in C([0,T],\mathbb{R}^d)$ and set $f_i = F(g_i), i = 1,2$. We have

$$|f_1(t) - f_2(t)| \le \int_0^t |b(f_1(s)) - b(f_2(s))|\,\mathrm{d}s + |g_1(t) - g_2(t)|.$$

Using the Lipschitz assumption and taking the supremum over time yields

$$\sup_{0\le r\le t} |f_1(r) - f_2(r)| \le \int_0^t L \sup_{0\le r\le s} |f_1(r) - f_2(r)|\,\mathrm{d}s + \|g_1 - g_2\|.$$

Applying Grönwall now gives

$$\|f_1 - f_2\| \leq e^{BT}\|g_1 - g_2\|.$$

Thus F is continuous, and in fact even Lipschitz continuous. Combining Theorem 5.3 and Theorem 5.4 gives that the family $\{X^\varepsilon\}_{\varepsilon>0}$ satisfies a LDP with (good) rate function

$$\tilde{I}(f) = \inf_{g\in\mathcal{H}_1 : F(g)=f} \frac{1}{2}\int_0^T |\dot{g}(s)| \, \mathrm{d}s.$$

We now wish to show that $\tilde{I} = I$. We first notice that F is a one-to-one mapping. Furthermore, for $g \in \mathcal{H}_1$ we have that $f - x \in \mathcal{H}_1$ and $\dot{f} = b(f) + \dot{g}$. Therefore we have that $\tilde{I} = I$, which completes the proof. ■

It turns out that the result above can be generalised, obtaining similar results for larger classes of SDEs under weaker assumptions. In particular we would like to present the following result from Baldi and Caramellino (2011). Let B be a m-dimensional Brownian motion. For $\varepsilon > 0$, consider the SDE

$$Y^\varepsilon(t) = b_\varepsilon(Y^\varepsilon(t))\mathrm{d}t + \varepsilon\sigma_\varepsilon(Y^\varepsilon(t))\mathrm{d}B(t), \quad Y^\varepsilon(0) = x, \tag{5.3}$$

with $b_\varepsilon : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma_\varepsilon : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ and $x \in \mathbb{R}^d$. We now make the following assumption.

Assumption 2. *There exist functions $b : \mathbb{R}^d \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ such that*

1. *for every $h \in \mathcal{H}_1([0,T], \mathbb{R}^m)$ and $x \in \mathbb{R}^d$ the ODE*

$$\dot{g}(t) = b(g(t)) + \sigma(g(t))\dot{h}(t), \quad g(0) = x \tag{5.4}$$

   has a unique solution on $[0,T]$.

2. *Let $S_x(h)$ denote the solution of Equation (5.4). For any $a > 0$, the restriction of S_X to the compact set $K_a = \{\|h\|_1 \leq a\}$ is continuous with respect to the uniform norm: for any $(h_n)_n \subset K_a$ such that $\|h_n - h\| \to 0$ with $h \in K_a$ then $\|S_x(h_n) - S_x(h)\| \to 0$. Here $\|\cdot\|$ denotes the uniform norm and $\|h\|_1$ is, for $h \in \mathcal{H}_1$, defined as*

$$\|h\|_1 = \left(\int_0^T |\dot{h}(s)|^2 \, \mathrm{d}s\right)^{\frac{1}{2}}.$$

3. *For every $R > 0, \varrho > 0, a > 0, c > 0$ there exist $\varepsilon_0 > 0, \alpha > 0$ such that, if $\varepsilon < \varepsilon_0$,*

$$\mathbb{P}\left(\|Y^\varepsilon - g\| > \varrho, \|\varepsilon B - h\| \leq \alpha\right) \leq e^{-\frac{R}{\varepsilon^2}}$$

   uniformly for $\|h\|_1 \leq a$ and $|x| \leq c$, where $g = S_x(h)$.

We now present the following theorem.

Theorem 5.6 (Theorem 2.4 in Baldi and Caramellino (2011))**.** *Suppose that $b_\varepsilon, \sigma_\varepsilon$ are locally Lipschitz and the SDE (5.3) has a strong solution for every $\varepsilon > 0$. Then, if Assumption 2 holds, the family $\{Y^\varepsilon\}_{\varepsilon>0}$ satisfies a LDP with inverse speed ε^2 and (good) rate function*

$$I_x(g) = \inf_{h\in\mathcal{H}_1 : S_x(h)=g} \frac{1}{2}\int_0^T |\dot{h}(s)|^2 \, \mathrm{d}s.$$

The assumptions of this theorem allows for a broad general class of SDEs. In particular, it can be shown that, when the functions b and σ are locally Lipschitz, when they satisfy a sub-linear growth condition and when $b_\varepsilon \to b$ and $\sigma_\varepsilon \to \sigma$ uniformly on compact sets, Assumption 2 is satisfied, and therefore Theorem 5.6 holds. From this we can conclude that the Freidlin-Wentzell theorem is indeed a special case of this theorem, with $b_\varepsilon = b$, $\sigma_\varepsilon = \sigma = 1$ and stricter assumptions.

We will leave the discussion of regular SDEs here. In the next section we will focus on a result describing a LDP for a class of SDEs that not only depend on their current position at time t, but on their entire trajectory on the interval $[0, t]$.

5.2. Large deviation principle for delay equations

Since the model we are considering, in particular the interacting particle system, not only depends on the current state but also at the state of the system in the past with a delay $\delta > 0$, we are interested in results on LDPs for classes of SDEs that allow such dependence. In this section we will present the results of Chiarini and Fischer (2014). They provide criteria under which a family of solutions $\{X^\varepsilon\}_{\varepsilon>0}$ of a SDE, depending not only on the current state of the system and its state in the past with a fixed delay, but possibly on its entire trajectory up to that point, satisfies a LDP. To be precise, we consider the following SDE

$$\mathrm{d}X^\varepsilon(t) = b_\varepsilon(t, X^\varepsilon)\mathrm{d}t + \sqrt{\varepsilon}\sigma_\varepsilon(t, X^\varepsilon)\mathrm{d}B(t), \tag{5.5}$$

and its controlled counterpart with control-parameter ν,

$$\mathrm{d}X^{\varepsilon,\nu}(t) = b_\varepsilon(t, X^{\varepsilon,\nu})\mathrm{d}t + \sigma_\varepsilon(t, X^{\varepsilon,\nu})\nu(t)\mathrm{d}t + \sqrt{\varepsilon}\sigma_\varepsilon(t, X^{\varepsilon,\nu})\mathrm{d}B(t), \tag{5.6}$$

with $X^\varepsilon(0) = X^{\varepsilon,\nu}(0) = x$ and $\nu \in \mathcal{M}^2([0, T], \mathbb{R}^m)$. Here b and b_ε are functions mapping $[0, T] \times C([0, T], \mathbb{R}^d)$ to $\mathbb{R}^d$, and σ and σ_ε functions mapping $[0, T] \times C([0, T], \mathbb{R}^d)$ to $\mathbb{R}^{d\times m}$, with m the dimension of the Brownian motion $B(t)$.

Also, for $f \in L^2([0, T], \mathbb{R}^m)$, consider the deterministic equation

$$\phi(t) = x + \int_0^t b(s, \phi)\,\mathrm{d}s + \int_0^t \sigma(s, \phi)f(s)\,\mathrm{d}s. \tag{5.7}$$

We are now ready to present the following set of assumptions, which will imply a LDP for the family $\{X^\varepsilon\}_{\varepsilon>0}$ satisfying (5.5).

Assumption 3. *H1 The coefficient b and σ are predictable. Moreover, $b(t, \cdot)$ and $\sigma(t, \cdot)$ are uniformly continuous on compact subsets of $C([0, T], \mathbb{R}^d)$, uniformly in $t \in [0, T]$, and $t \mapsto \sigma(t, \phi)$ is in $L^2([0, T], \mathbb{R}^{d\times m})$ for any $\phi \in C([0, T], \mathbb{R}^d)$.*

H2 The coefficients b_ε and σ_ε are predictable maps such that $b_\varepsilon \to b$ and $\sigma_\varepsilon \to \sigma$ as $\varepsilon \to 0$ uniformly on $[0, T] \times C([0, T], \mathbb{R}^d)$.

H3 For all sufficiently small $\varepsilon > 0$, pathwise uniqueness and existence in the strong sense hold for (5.5).

H4 For any $f \in L^2([0, T], \mathbb{R}^m)$, (5.7) *has a unique solution so that the map*

$$\Gamma_x : L^2([0, T], \mathbb{R}^m) \to C([0, T], \mathbb{R}^d)$$

which takes $f \in L^2([0, T], \mathbb{R}^m)$ to the solution of (5.7) *is well defined.*

H5 For all $N \in \mathbb{N}$, the map Γ_x is continuous when restricted to

$$S_N := \left\{ f \in L^2([0,T], \mathbb{R}^m) : \int_0^T |f(s)|^2 \, \mathrm{d}s \leq N \right\}$$

endowed with the weak topology of $L^2([0,T], \mathbb{R}^m)$.

H6 If $\{\varepsilon_n\}_n \subset (0,1]$ is such that $\varepsilon_n \to 0$ as $n \to \infty$ and $\{\nu_n\}_n \subset \mathcal{M}^2([0,T], \mathbb{R}^m)$ is such that, for some constant $N > 0$,

$$\sup_{n \in \mathbb{N}} \int_0^T |\nu^n(s, \omega)|^2 \, \mathrm{d}s \leq N \quad \text{for } \theta\text{-almost all } \omega,$$

then $\{X^{\varepsilon_n, \nu_n}\}_n$ is tight as a family of $C([0,T], \mathbb{R}^d)$-valued random variables and

$$\sup_{n \in \mathbb{N}} \int_0^T \mathbb{E}\left[|\sigma(s, X^{\varepsilon_n, \nu_n})|^2 \right] \, \mathrm{d}s < \infty.$$

As stated, under these assumptions we have that the family $\{X^\varepsilon\}_{\varepsilon>0}$ satisfies a LDP. To be precise, we have the following theorem.

Theorem 5.7. *Assume that Assumption 3 holds. Then the family $\{X^\varepsilon\}_{\varepsilon>0}$ of solutions to the SDE* (5.5) *with initial condition $X^\varepsilon(0) = x$ satisfies a LDP with good rate function*

$$I_x(\phi) = \inf_{\{f \in L^2([0,T], \mathbb{R}^m) : \Gamma_x(f) = \phi\}} \frac{1}{2} \int_0^T |f(t)|^2 \, \mathrm{d}t,$$

with the convention that $\inf\{\varnothing\} = +\infty$.

Although the result is powerful and allows for a broad class of SDEs, the assumptions are also not easy to work with. We will also present a second set of assumptions which will imply Assumption 3. While these assumptions are stricter than the ones given above, they are also easier to check. These assumptions are as follows.

Assumption 4. *A1 The functions b and σ satisfy a sub-linear growth condition. Specifically, there exists $M > 0$ such that for all $t \in [0,T]$ and all $\phi \in C([0,T], \mathbb{R}^d)$*

$$|b(t, \phi)| \wedge |\sigma(t, \phi)| \leq M \left(1 + \sup_{0 \leq s \leq t} |\phi(s)| \right).$$

A2 The functions b and σ are locally Lipschitz. Specifically, for any $R > 0$, there exists $L_R > 0$ such that, for all $t \in [0,T]$ and all $\phi, \tilde{\phi} \in C([0,T], \mathbb{R}^d)$ with $sup_{0 \leq s \leq t} |\phi(s)| \wedge |\tilde{\phi}(s)| \leq R$

$$|b(t, \phi) - b(t, \tilde{\phi})| \wedge |\sigma(t, \phi) - \sigma(t, \tilde{\phi})| \leq L_R \sup_{0 \leq s \leq t} |\phi(s) - \tilde{\phi(s)}|.$$

A3 The functions b_ε and σ_ε enjoy property A1 with the same constant M as well as property A2.

A4 The coefficients b_ε and σ_ε converge as $\varepsilon \to 0$ to b and σ respectively, uniformly on bounded subsets of $[0,T] \times C([0,T], \mathbb{R}^d)$.

The proof that Assumption 4 implies Assumption 3 is structured as follows. Assumption H1 is implied by A2 and A1. Assumption H3 is implied by A3. H4 and H5 can be proven by assuming A2 and A1. Assumption H6 can be proven with A1 and the Kolmogorov tightness criterion. Lastly, A4 is a weaker version of H2, but in combination with A3 it can be shown that the assumptions suffice such that Theorem 5.7 is also true when Assumption 4 holds. The details can be found in Chiarini and Fischer (2014).

Remark. *If $b(t,\phi) = \tilde{b}(t,\phi(t))$, $\sigma(t,\phi) = \tilde{\sigma}(t,\phi(t))$, i.e. the functions only depend on the current state and not on their past trajectory, and $\tilde{\sigma}$ is a square matrix such that $\tilde{\sigma}(t,x)\tilde{\sigma}(t,x)^T$ is uniformly positive definite, then the rate function can be expressed as*

$$I_x(\phi) = \frac{1}{2}\int_0^T \left(\dot{\phi}(s) - \tilde{b}(s,\phi(s))\right)^T \left(\tilde{\sigma}(t,x)\tilde{\sigma}(t,x)^T\right)^{-1} \left(\dot{\phi}(s) - \tilde{b}(s,\phi(s))\right) \mathrm{d}s$$

whenever $\phi - x \in \mathcal{H}_1([0,T],\mathbb{R}^d)$ and $I_x(\phi) = +\infty$ otherwise.

With this powerful theorem in hands, we now wish to see if we can apply it, or one of the previous mentioned theorems, to our model. In the next section we investigate which assumptions we need to make on our model to show that a LDP holds for the solutions of the SDEs we have been studying so far.

5.3. Application to our model

Now that we have studied various theorems about LDPs for solutions of SDEs, we wish to see when and how they can be applied to the two equations we have been studying so far. We will consider the two equations separately. We start with the Mckean-Vlasov equation. Let $T, \delta, \varepsilon > 0$, consider the SDE

$$\begin{aligned} \mathrm{d}X^\varepsilon(t) &= f(X^\varepsilon(t))\mathrm{d}t + g(\mathbb{E}\left[X^\varepsilon(t-\delta)\right])\mathrm{d}t + \varepsilon\mathrm{d}B(t), \\ X^\varepsilon(0) &= x, \\ \mathbb{E}\left[X^\varepsilon(s)\right] &= x(s), \quad s \in [-\delta, 0). \end{aligned} \tag{5.8}$$

Assume that Assumption 1 holds, such that we know that solutions exist for all $\varepsilon > 0$. Now define $h_\varepsilon(t) = g(\mathbb{E}\left[X^\varepsilon(t-\delta)\right])$, which is a deterministic function from $[0,T]$ to $\mathbb{R}^d$. Now also define $b_\varepsilon(t, X^\varepsilon(t)) = f(X^\varepsilon(t)) + h_\varepsilon(t)$. We can now rewrite our equation to

$$\begin{aligned} \mathrm{d}X^\varepsilon(t) &= b_\varepsilon(t, X^\varepsilon(t))\mathrm{d}t + \varepsilon\mathrm{d}B(t), \\ X^\varepsilon(0) &= x. \end{aligned}$$

This is an equation of the form of (5.5). We therefore have that this equation satisfies a LDP when Assumption 4 is met. Since we have that $\sigma = \sigma_\varepsilon = 1$, they clearly satisfy a sub-linear growth condition and are locally Lipschitz. Furthermore, we trivially have that $\sigma_\varepsilon \to \sigma$ as $\varepsilon \to 0$. All that is left to show is that the same holds for our functions b and b_ε. First we need to define our function b. We define $b(x,t) = f(x) + g(\varphi(t-\delta))$, where φ satisfies the ODE

$$\begin{aligned} \varphi(t) &= x + \int_0^t f(\varphi(s)) + g(\varphi(s-\delta))\,\mathrm{d}s, \quad t \in [0,T], \\ \varphi(0) &= x, \varphi(t) = x(t), \quad t \in [-\delta, 0). \end{aligned} \tag{5.9}$$

By Theorem 3.4, with $\varepsilon = 0$, we have that φ is uniformly bounded on $[-\delta, T]$. Since we already assumed g to be Lipschitz, we also have that $g(\varphi(\cdot - \delta))$ is uniformly bounded. If we now assume that f satisfies a sub-linear growth condition we find that b also satisfies a sub-linear growth condition. Equivalently, since we have that $\mathbb{E}\left[X^\varepsilon(\cdot - \delta)\right]$ is again uniformly bounded by Theorem 3.4, we also find that b_ε satisfies a sub-linear growth condition if f satisfies such condition. If we assume that $0 < \varepsilon \leq M$ for some $M > 0$, we can also make the constant in the growth condition independent of ε by taking the supremum over all $\varepsilon \in [0, M]$, which is a finite number.

The local Lipschitz assumption also holds. Since we have, by Assumption 1, that f is locally Lipschitz, we also have that b and b_ε are locally Lipschitz with the same constant as f.

For the final condition we need to show that $b_\varepsilon \to b$ uniformly on bounded subsets of $[0,T] \times C([0,T], \mathbb{R}^d)$. This is equivalent to showing that $g(\mathbb{E}[X^\varepsilon(\cdot - \delta)]) \to g(\varphi)$ uniformly on $[0,T]$. By the Lipschitz assumption of g this is implied if $\mathbb{E}[X^\varepsilon(\cdot - \delta)] \to \varphi$ uniformly on $[0,T]$. The results of Chapter 3, in particular Theorem 3.4 is not sufficient to show this result. We will now derive a series of results that will allow us to show that the expectation of the solution of (5.8) indeed converges uniformly to the solution of (5.9).

5.3.1. Central moments

Consider the SDE (5.8). Assume that Assumption 1 holds. Furthermore, assume that f is globally Lipschitz. We obtain the following result.

Lemma 5.8. *There exists a constant C_2 such that for all $t \in [0,T]$ we have*

$$\mathbb{E}\left[\sup_{0\le u\le t} |X^\varepsilon(u) - \mathbb{E}[X^\varepsilon(u)]|^2\right] \le e^{(1+L^2)t}(td\varepsilon^2 + C_2\sqrt{t}\varepsilon).$$

In particular, when $\varepsilon \to 0$ we have that the second central moment also goes to 0.

Proof. We proof this lemma by doing a straight forward computation, using our assumptions and Grönwall's lemma to prove the inequality. In particular, using Itô's formula we have

$$\begin{aligned}|X^\varepsilon(t) - \mathbb{E}[X^\varepsilon(t)]|^2 = &\int_0^t 2\langle X^\varepsilon(s) - \mathbb{E}[X^\varepsilon(s)], f(X^\varepsilon(s)) + \mathbb{E}[f(X^\varepsilon(s))]\rangle \,\mathrm{d}s \\ &+ td\varepsilon^2 + \int_0^t 2\varepsilon\langle X^\varepsilon(s), \mathrm{d}B(s)\rangle.\end{aligned}$$

Applying Cauchy-Schwarz and Young's inequality to the first inner product yields

$$\begin{aligned}|X^\varepsilon(t) - \mathbb{E}[X^\varepsilon(t)]|^2 = &\int_0^t |X^\varepsilon(s) - \mathbb{E}[X^\varepsilon(s)]|^2 + |f(X^\varepsilon(s)) - \mathbb{E}[f(X^\varepsilon(s))]|^2 \,\mathrm{d}s \\ &+ td\varepsilon^2 + \int_0^t 2\varepsilon\langle X^\varepsilon(s), \mathrm{d}B(s)\rangle.\end{aligned}$$

Taking supremum over time and expectation gives us

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le u\le t} |X^\varepsilon(u) - \mathbb{E}[X^\varepsilon(u)]|^2\right] \le &\int_0^t \mathbb{E}\left[|X^\varepsilon(s) - \mathbb{E}[X^\varepsilon(s)]|^2\right] \mathrm{d}s \\ &+ \int_0^t \mathbb{E}\left[|f(X^\varepsilon(s)) - \mathbb{E}[f(X^\varepsilon(s))]|^2\right] \mathrm{d}s \\ &+ td\varepsilon^2 + \mathbb{E}\left[\sup_{0\le u\le t}\int_0^u 2\varepsilon\langle X^\varepsilon(s), \mathrm{d}B(s)\rangle\right].\end{aligned}$$

Using the fact that $\mathbb{E}[|Y - \mathbb{E}[Y]|^2] \le \mathbb{E}[|]Y - a|^2$ for any a, with $Y = f(X^\varepsilon(s))$ and $a = f(\mathbb{E}[X^\varepsilon(s)])$ we get

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le u\le t} |X^\varepsilon(u) - \mathbb{E}[X^\varepsilon(u)]|^2\right] \le &\int_0^t \mathbb{E}\left[|X^\varepsilon(s) - \mathbb{E}[X^\varepsilon(s)]|^2\right] \mathrm{d}s \\ &+ \int_0^t \mathbb{E}[|f(X^\varepsilon(s)) - f(\mathbb{E}[X^\varepsilon(s)])|^2 \,\mathrm{d}s \\ &+ td\varepsilon^2 + \mathbb{E}\left[\sup_{0\le u\le t}\int_0^u 2\varepsilon\langle X^\varepsilon(s), \mathrm{d}B(s)\rangle\right].\end{aligned}$$

Using the Lipschitz property and the The Burkholder-Davis-Gundy inequality yields

$$\begin{aligned}
&\leq \int_0^t \mathbb{E}\left[|X^\varepsilon(s) - \mathbb{E}\left[X^\varepsilon(s)\right]|^2\right] \, \mathrm{d}s \\
&\quad + L^2\mathbb{E}\left[|(X^\varepsilon(s) - \mathbb{E}\left[(X^\varepsilon(s)\right]|^2\right] \, \mathrm{d}s \\
&\qquad + td\varepsilon^2 + C_1\mathbb{E}\left[\left(\int_0^t \varepsilon^2|X^\varepsilon(s)|^2 \, \mathrm{d}s\right)^{\frac{1}{2}}\right].
\end{aligned}$$

Using Jensen's inequality and the fact that the second moment of X^ε is bounded we can bound the last term by $C_2\sqrt{t}\varepsilon$ for some $C_2 > 0$ independent of ε when $\varepsilon \in [0, M]$. Finally, applying Grönwall's lemma yields

$$\mathbb{E}\left[\sup_{0\leq u\leq t} |X^\varepsilon(u) - \mathbb{E}\left[X^\varepsilon(u)\right]|^2\right] \leq e^{(1+L^2)t}(td\varepsilon^2 + C_2\sqrt{t}\varepsilon),$$

which proves the lemma. ■

We now also define a new process Y^ε, which will help us show that the mean of X^ε converges towards φ, the solution of the corresponding ODE. We define Y^ε as the solution of the SDE

$$\begin{aligned}
\mathrm{d}Y^\varepsilon(t) &= f(Y^\varepsilon(t))\mathrm{d}t + g(Y^\varepsilon(t-\delta)) + \varepsilon \mathrm{d}B(t), \\
Y^\varepsilon(0) &= x, \\
Y^\varepsilon(s) &= x(s), \quad s \in [-\delta, 0),
\end{aligned}$$

where the Brownian motion B is the same as for the process X^ε. Notice that the equation for Y^ε coincides with the interacting particle system with $N = 1$. Therefore we have that Theorem 3.1 and Theorem 3.2 also apply to the process Y^ε. In particular, strong solutions exist and those solutions have finite moments. We now have the following lemma.

Lemma 5.9. *For all $t \in [0, T]$ we have*

$$\mathbb{E}\left[\sup_{0\leq u\leq t} |X^\varepsilon(u) - Y^\varepsilon(u)|^2\right] \leq e^{(1+2C+2L+L^2)T}(\frac{1}{2}t^2d\varepsilon^2 + \frac{2}{3}C_2t^{\frac{3}{2}}\varepsilon),$$

where C_2 is the same constant as in Lemma 5.8.

Proof. The proof is similar to that of Lemma 5.8. Again, using Ito's formula we compute

$$\begin{aligned}
|X^\varepsilon(t) - Y^\varepsilon(t)|^2 &= \int_0^t 2\langle X^\varepsilon(s) - Y^\varepsilon(s), f(X^\varepsilon(s)) - f(Y^\varepsilon(s))\rangle \, \mathrm{d}s \\
&\quad + \int_0^t 2\langle X^\varepsilon(s) - Y^\varepsilon(s), g(\mathbb{E}\left[X^\varepsilon(s-\delta)\right] - g(Y^\varepsilon(s-\delta)))\, \mathrm{d}s.
\end{aligned}$$

For the first inner-product we use the one-sided Lipschitz assumption. For the second inner-product we apply Cauchy-Schwarz, the global Lipschitz assumption on g and Young's inequality. Together this yields

$$\begin{aligned}
&\leq \int_0^t 2C|X^\varepsilon(s) - Y^\varepsilon(s)|^2 + L|X^\varepsilon(s) - Y^\varepsilon(s)|^2 \, \mathrm{d}s \\
&\quad + \int_0^t L|\mathbb{E}\left[X^\varepsilon(s-\delta)\right] - Y^\varepsilon(s-\delta)|^2 \, \mathrm{d}s.
\end{aligned}$$

Adding and subtracting the term $X^\varepsilon(s-\delta)$ and using the triangle inequality gives us

$$\begin{aligned}
&\leq \int_0^t (2C+L)|X^\varepsilon(s)-Y^\varepsilon(s)|^2 + L|\mathbb{E}\left[X^\varepsilon(s-\delta)\right] - X^\varepsilon(s-\delta)|^2 \,\mathrm{d}s \\
&\quad + \int_0^t L|X^\varepsilon(s-\delta)-Y^\varepsilon(s-\delta)|^2 \,\mathrm{d}s.
\end{aligned}$$

Taking supremum over time and expectation yields

$$\begin{aligned}
\mathbb{E}\left[\sup_{0\leq u\leq t}|X^\varepsilon(u)-Y^\varepsilon(u)|^2\right] &\leq \int_0^t (2C+2L)\mathbb{E}\left[\sup_{0\leq u\leq s}|X^\varepsilon(u)-Y^\varepsilon(u)|^2\right] \\
&\quad + \int_0^t L\mathbb{E}\left[|X^\varepsilon(s-\delta)-\mathbb{E}\left[X^\varepsilon(s-\delta)\right]|^2\right] \,\mathrm{d}s.
\end{aligned}$$

Using Grönwall's lemma we get

$$\mathbb{E}\left[\sup_{0\leq u\leq t}|X^\varepsilon(u)-Y^\varepsilon(u)|^2\right] \leq e^{(2C+2L)t}\int_0^t L\mathbb{E}\left[|X^\varepsilon(s-\delta)-\mathbb{E}\left[X^\varepsilon(s-\delta)\right]|^2\right] \,\mathrm{d}s.$$

Finally, using Lemma 5.8

$$\begin{aligned}
&\leq e^{(2C+2L)t}\int_0^t e^{(1+L^2)T}(sd\varepsilon^2 + C_2\sqrt{s}\varepsilon)\,\mathrm{d}s \\
&= e^{(1+2C+2L+L^2)T}(\frac{1}{2}t^2 d\varepsilon^2 + \frac{2}{3}C_2 t^{\frac{3}{2}}\varepsilon),
\end{aligned}$$

which proves the lemma. ■

Now instead of assuming that f is globally Lipschitz, assume that Assumption 1 holds and that f can be approximated by functions $f_R, R>0$ such that $f=f_R$ on $B(0,R)$, f_R satisfies Assumption 1 and f_R is globally Lipschitz with Lipschitz coefficient L_R such that there exists $L>0$ such that $L_R \leq LR$ for all $R>0$. We have the following lemma regarding the second moment.

Lemma 5.10. *There exist $C_1, C_2 > 0$ such that for all $t\in[0,T]$*

$$\mathbb{E}\left[\sup_{0\leq u\leq t}|X^\varepsilon(u)-\mathbb{E}\left[X^\varepsilon(u)\right]|^2\right] \leq 6C_1\sqrt{\frac{2L^2T}{\ln\frac{1}{\varepsilon}}} + 3e^T(td\varepsilon^{\frac{3}{2}} + C_2\sqrt{t\varepsilon}).$$

Here the constant C_2 is the same as in Lemma 5.8 and the constant C_1 can be chosen independent of ε when $\varepsilon\in[0,M]$ for some $M>0$. In particular, the second central moment of X^ε in the supremum norm tends to 0 when $\varepsilon\to 0$.

Proof. The proof works by approximating the solution X^ε with other processes which do satisfy the assumptions of Lemma 1. In particular, in accordance with our assumptions, let f_R be such that it is globally Lipschitz with Lipschitz constant L_R, that it coincides with f on $B(0,R)$ and that it satisfies Assumption 1. Let $L>0$ such that $L_R\leq LR$ for all $R>0$. Let X_R^ε satisfy the SDE

$$\begin{aligned}
\mathrm{d}X_R^\varepsilon(t) &= f(X_R^\varepsilon(t))\mathrm{d}t + g(\mathbb{E}\left[X_R^\varepsilon(t-\delta)\right])\mathrm{d}t + \varepsilon\mathrm{d}B(t), \\
X_R^\varepsilon(0) &= x, \\
X_R^\varepsilon(s) &= x(s), \quad s\in[-\delta,0),
\end{aligned}$$

where B denotes the same Brownian motion as that of the process X^ε. We have that

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le u\le t}|X^\varepsilon(u)-\mathbb{E}\left[X^\varepsilon(u)\right]|^2\right] &\le 3\mathbb{E}\left[\sup_{0\le u\le t}|X^\varepsilon(u)-X_R^\varepsilon(u)|^2\right]\\ &\quad+3\mathbb{E}\left[\sup_{0\le u\le t}|X_R^\varepsilon(u)-\mathbb{E}\left[X_R^\varepsilon(u)\right]|^2\right]\\ &\quad+3\mathbb{E}\left[\sup_{0\le u\le t}|\mathbb{E}\left[X_R^\varepsilon(u)\right]-\mathbb{E}\left[X^\varepsilon(u)\right]|^2\right].\end{aligned}$$

For the last term we notice that

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le u\le t}|\mathbb{E}\left[X_R^\varepsilon(u)\right]-\mathbb{E}\left[X^\varepsilon(u)\right]|^2\right] &= \sup_{0\le u\le t}|\mathbb{E}\left[X_R^\varepsilon(u)-X^\varepsilon(u)\right]|^2\\ &\le \mathbb{E}\left[\sup_{0\le u\le t}|X_R^\varepsilon(u)-X^\varepsilon(u)|^2\right].\end{aligned}$$

For the second term Lemma 5.8 applies such that it is bounded by

$$e^{(1+L_R^2)t}(td\varepsilon^2+C_2\sqrt{t}\varepsilon),$$

where C_2 can be chosen independent of R. We now wish to control the first term (which, as shown above, also serves as a bound for the third term). Using Theorem 2.19 we have that

$$\mathbb{E}\left[\sup_{0\le u\le t}|X^\varepsilon(u)-X_R^\varepsilon(u)|^2\right]=\mathbb{E}\left[\sup_{0\le u\le t}|X^\varepsilon(u)-X_R^\varepsilon(u)|^2\mathbb{1}\{\sup_{0\le u\le t}X^\varepsilon(u)>R\}\right].$$

Using Hölder's theorem with $p=q=2$ we get

$$\le\left(\mathbb{E}\left[\sup_{0\le u\le t}|X^\varepsilon(u)-X_R^\varepsilon(u)|^4\right]\right)^{\frac{1}{2}}\left(\mathbb{P}\left(\tau_R<t\right)\right)^{\frac{1}{2}},$$

where τ_R denotes the exit time of the ball $B(0,R)$ of the process X^ε. Using that $|a-b|^4\le 8|a|^4+8|b|^4$ for all $a,b\in\mathbb{R}^d$ we obtain

$$\le\left(\mathbb{E}\left[\sup_{0\le u\le t}8|X^\varepsilon(u)|^4\right]+\mathbb{E}\left[\sup_{0\le u\le t}8|X_R^\varepsilon(u)|^4\right]\right)^{\frac{1}{2}}\left(\mathbb{P}\left(\tau_R<t\right)\right)^{\frac{1}{2}}.$$

Using Theorem 3.4 we have that there exists a constant $\tilde{C}_1$, independent of R and ε for $\varepsilon\in[0,M]$ for some $M>0$ which bounds the first term. For the second term we invoke Equation (3.9) with $p=2$. Together this yields

$$\mathbb{E}\left[\sup_{0\le u\le t}|X^\varepsilon(u)-X_R^\varepsilon(u)|^2\right]\le\tilde{C}_1\sqrt{(1+|x|^2)e^{KT}}\frac{1}{R}=C_1\frac{1}{R}.$$

Combining the three terms together, and using the assumption that $L_R\le LR$, we obtain

$$\mathbb{E}\left[\sup_{0\le u\le t}|X^\varepsilon(u)-\mathbb{E}\left[X^\varepsilon(u)\right]|^2\right]\le 6C_1\frac{1}{R}+3e^{(1+L^2R^2)T}(td\varepsilon^2+C_2\sqrt{t}\varepsilon).$$

Taking $R=\sqrt{\frac{1}{2L^2T}\ln\frac{1}{\varepsilon}}$, we get

$$\le 6C_1\sqrt{\frac{2L^2T}{\ln\frac{1}{\varepsilon}}}+3e^T(td\varepsilon^{\frac{3}{2}}+C_2\sqrt{t\varepsilon}),$$

which proves the theorem. ■

Remark. *For the lemma above we made the assumption that $L_R \leq LR$ for some $L > 0$. Instead, we could also assume that there exist $L, \alpha > 0$ such that $L_R \leq LR^\alpha$. Under this assumption we can derive a similar result, where we also have that the second central moment tends to 0 when $\varepsilon \to 0$. The only difference is that we have to change the rate at which we send $R \to \infty$ when we send $\varepsilon \to 0$. In particular, taking*

$$R = \left(\frac{1}{2L^2 T} \ln \frac{1}{\varepsilon} \right)^{\frac{1}{2\alpha}}$$

would suffice to show the result.

Before we continue we would like to remark that under Assumption 1, without the global Lipschitz assumption on f we still have that

$$\mathbb{E}\left[\sup_{0 \leq u \leq t} |X^\varepsilon(u) - Y^\varepsilon(u)|^2 \right] \to 0,$$

as $\varepsilon \to 0$. Indeed, closer inspection of the proof of Lemma 5.9 only requires the global Lipschitz assumption to invoke Lemma 5.8. Now that we can replace that result by the result of Lemma 5.10, we get a similar result which still tends to 0 when $\varepsilon \to 0$.

5.3.2. Large deviation principles (LDPs)

Now that we have analysed how the second central moment of X^ε behaves in terms of ε, we can return to our problem where we try to find whether or not a LDP holds. As explained at the beginning of this section, under the assumption that Assumption 1 holds and that f satisfies a sub-linear growth condition, we only needed to show that $\mathbb{E}[X^\varepsilon(t)] \to \varphi(t)$ uniformly on $[0, T]$. Assume that, for the Lipschitz coefficient L_R of Assumption 1, it holds that there exists a $\alpha > 0$, such that $L_R \leq LR^\alpha$ for all $R > 0$. We now have the following lemma.

Lemma 5.11. *Under the assumptions mentioned above, we have*

$$\sup_{0 \leq t \leq T} |\mathbb{E}[X^\varepsilon(t)] - \varphi(t)|^2 \to 0,$$

as $\varepsilon \to 0$. Here φ is defined as the solution of (5.9).

Proof. To proof this limit we make use of the lemmas presented in the previous subsection. In particular, we use that

$$\begin{aligned}\sup_{0 \leq t \leq T} |\mathbb{E}[X^\varepsilon(t)] - \varphi(t)|^2 \leq 3\mathbb{E}\left[\sup_{0 \leq t \leq T} |\mathbb{E}[X^\varepsilon(t)] - X^\varepsilon(t)|^2 \right] + 3\mathbb{E}\left[\sup_{0 \leq t \leq T} |X^\varepsilon(t) - Y^\varepsilon(t)|^2 \right] \\ + 3\mathbb{E}\left[\sup_{0 \leq t \leq T} |Y^\varepsilon(t) - \varphi(t)|^2 \right]. \qquad (5.10)\end{aligned}$$

In Lemma 5.10 and Lemma 5.9 we have seen that the first and second term tend to 0 when $\varepsilon \to 0$. For the last term we carry out the following computation. By Itô's formula we have

$$\begin{aligned}|Y^\varepsilon(t) - \varphi(t)|^2 &= \int_0^t 2\langle Y^\varepsilon(s) - \varphi(s), f(Y^\varepsilon(s)) - f(\varphi(s))\rangle \,\mathrm{d}s \\ &+ \int_0^t 2\langle Y^\varepsilon(s) - \varphi(s), g(Y^\varepsilon(s-\delta)) - g(\varphi(s-\delta))\rangle \,\mathrm{d}s + td\varepsilon^2 \\ &+ \int_0^t \varepsilon\langle Y^\varepsilon(s), \mathrm{d}B(s)\rangle.\end{aligned}$$

Applying the one-sided Lipschitz assumption to the first inner-product and Cauchy-Schwarz, the global Lipschitz assumption and Young's inequality to the second inner-product yields

$$\begin{aligned}\leq \int_0^t 2C|Y^\varepsilon(s)-\varphi(s)|^2 + L|Y^\varepsilon(s)-\varphi(s)|^2\,\mathrm{d}s \\ + \int_0^t L|Y^\varepsilon(s-\delta)-\varphi(s-\delta)|^2\,\mathrm{d}s + td\varepsilon^2 + \int_0^t 2\varepsilon\langle Y^\varepsilon(s),\mathrm{d}B(s)\rangle.\end{aligned}$$

Taking supremum over time and expectation gives

$$\begin{aligned}\mathbb{E}\left[\sup_{0\leq u\leq t}|Y^\varepsilon(u)-\varphi(u)|^2\right] \leq \int_0^t (2C+2L)\mathbb{E}\left[\sup_{0\leq u\leq s}|Y^\varepsilon(u)-\varphi(u)|\right]\mathrm{d}s \\ + td\varepsilon^2 + \mathbb{E}\left[\sup_{0\leq u\leq t}\int_0^u 2\varepsilon\langle Y^\varepsilon(s),\mathrm{d}B(s)\rangle\right].\end{aligned}$$

Using Burkholder-Davis-Gundy Inequality gives

$$\begin{aligned}\leq \int_0^t (2C+2L)\mathbb{E}\left[\sup_{0\leq u\leq s}|Y^\varepsilon(u)-\varphi(u)|\right]\mathrm{d}s \\ + td\varepsilon^2 + C_2\mathbb{E}\left[\left(\int_0^t 4\varepsilon^2|Y^\varepsilon(s)|^2\,\mathrm{d}s\right)^{\frac{1}{2}}\right].\end{aligned}$$

We now apply Grönwall's inequality, which gives us

$$\mathbb{E}\left[\sup_{0\leq t\leq T}|Y^\varepsilon-\varphi(t)|^2\right] \leq e^{(2C+2L)T}\left(Td\varepsilon^2 + \varepsilon C_2\mathbb{E}\left[\left(\int_0^t |Y^\varepsilon s|^2\,\mathrm{d}s\right)^{\frac{1}{2}}\right]\right).$$

Since the second moment of Y^ε is uniformly bounded in ε for ε bounded, we have that the right hand side converges to 0 when $\varepsilon \to 0$. Therefore we have that all terms in the right hand side of (5.10) tend to 0 when $\varepsilon \to 0$, proving the lemma. ■

The above lemma directly gives rise to the following theorem, describing a LDP for the family $\{X^\varepsilon\}_{\varepsilon>0}$ using Theorem 5.7.

Theorem 5.12. *Consider the SDE* (5.8). *Assume that Assumption 1 holds and that f satisfies a sub-linear growth condition. Furthermore assume that for the Lipschitz coefficient L_R of Assumption 1 there exists an $\alpha > 0$ such that $L_R \leq LR^\alpha$ for all $R > 0$. Then the family $\{X^\varepsilon\}_{\varepsilon>0}$ satisfies a LDP with inverse speed ε^2 and (good) rate function*

$$I_x(\phi) = \frac{1}{2}\int_0^T |\dot{\phi}(t) - f(\phi(t)) - g(\varphi(t-\delta))|^2\,\mathrm{d}t$$

for $\phi - x \in \mathcal{H}_1([0,T],\mathbb{R}^d)$ and $I_x(\phi) = +\infty$ else. Here φ is the solution of (5.9).

The proof follows by combining the arguments given above. The rate function can be found by realising that the map Γ, defined in Assumption 3, is invertible, which allows for an explicit expression of the rate function. We will end our discussion of the McKean-Vlasov equation for now. It is possible that under milder assumptions a similar theorem can be shown, by showing that those assumptions imply Assumption 3 directly.

We will now discuss the interacting-particle system. Let $N \geq 1$, $T, \delta > 0$ fixed. Consider the system of SDEs

$$\begin{aligned}&\mathrm{d}X^{i,\varepsilon}(t) = f(X^{i,\varepsilon}(t))\mathrm{d}t + g\left(\frac{1}{N}\sum_{j=1}^N X^{j,\varepsilon}(t-\delta)\right)\mathrm{d}t + \varepsilon\mathrm{d}B^i(t),\\ &X^{i,\varepsilon}(0) = x^i,\\ &X^{i,\varepsilon}(s) = x^i(s), s\in[-\delta,0), i = 1,\dots,N.\end{aligned}$$

Here B^i are independent Brownian motions. Equivalently, we can consider the vectorized equation

$$\mathrm{d}X^\varepsilon(t) = F(X^\varepsilon(t))\mathrm{d}t + G(X^\varepsilon(t-\delta))\mathrm{d}t + \varepsilon \mathrm{d}B(t), \tag{5.11}$$

with

$$F(X^\varepsilon(t)) = \begin{pmatrix} f(X^{1,\varepsilon}(t)) \\ f(X^{2,\varepsilon}(t)) \\ \vdots \\ f(X^{N,\varepsilon}(t)) \end{pmatrix}, \quad G(X^\varepsilon(t-\delta)) = \begin{pmatrix} g\left(\frac{1}{N}\sum_{j=1}^N X^{j,\varepsilon}(t-\delta)\right) \\ g\left(\frac{1}{N}\sum_{j=1}^N X^{j,\varepsilon}(t-\delta)\right) \\ \vdots \\ g\left(\frac{1}{N}\sum_{j=1}^N X^{j,\varepsilon}(t-\delta)\right) \end{pmatrix}, \quad B(t) = \begin{pmatrix} B^1(t) \\ B^2(t) \\ \vdots \\ B^N(t) \end{pmatrix}$$

and equivalently for the initial conditions. Now Equation (5.11) is already in the form of (5.5) with

$$b_\varepsilon(X^\varepsilon, t) = b(X^\varepsilon, t) = F(X^\varepsilon(t)) + G(X^\varepsilon(t-\delta))$$

and $\sigma_\varepsilon = \sigma = 1$. Checking the assumptions of Assumption 4 we find that point A4 trivially holds. Furthermore, point A3 holds if A2 and A1 hold, as the functions b_ε and σ_ε do not depend on ε. Since we have that F and G are locally Lipschitz and satisfy a sub-linear growth condition if and only if their components do, we obtain the following theorem.

Theorem 5.13. *Consider the SDE* (5.11). *Assume that the functions f and g are locally Lipschitz and satisfy a sub-linear growth condition. Then the family $\{X^\varepsilon\}_{\varepsilon>0}$ satisfies a LDP with inverse speed ε^2 and (good) rate function*

$$I_x(\phi) = \frac{1}{2}\int_0^T |\dot{\phi}(t) - F(\phi(t)) - G(\phi(t-\delta))|^2 \,\mathrm{d}t$$

for $\phi - x \in \mathcal{H}_1([0,T], \mathbb{R}^{dN})$ and $I_x(\phi) = +\infty$ else.

Remark. *In the above, both in Theorem 5.12 and in Theorem 5.13, when $t \in [0,\delta)$, we need an interpretation of the term $g(\phi(t-\delta))$ for functions $\phi : [0,T] \to \mathbb{R}^d$. One option is to extend the domain of the function ϕ to $[-\delta, T]$ and define $\phi(t) = x(t)$ for $t \in [-\delta, 0)$, where $x(\cdot)$ is the initial condition of the SDE in question. Another option is to redefine the function g, i.e.*

$$g(\phi(t-\delta)) = g(\phi(t-\delta), t) = \begin{cases} g(x(t-\delta)), & t \in [0,\delta), \\ g(\phi(t-\delta)), & t \in [\delta, T]. \end{cases}$$

Although the rate functions for the McKean-Vlasov process and the interacting particle process are different, we would like to note that they have the same expected asymptotic behaviour. The expected behaviour of the interacting particle process minimizes its rate function. The function ϕ that minimizes the rate function satisfies

$$\dot{\phi} - F(\phi(t)) - G(\phi(t-\delta)) = 0.$$

If we assume that the initial conditions are identical for all particles, we have

$$\frac{1}{N}\sum_{j=1}^N \phi^j(t) = \phi^i(t)$$

by symmetry of the equations. This means that we can decouple the equations and solve them component-wise. We thus have that the component ϕ^i satisfies the ODE

$$\frac{\mathrm{d}\phi^i(t)}{\mathrm{d}t} = f(\phi^i(t)) + g(\phi(t-\delta)).$$

This is precisely the same ODE as φ solves. In Lemma 5.11 we have already seen that this is the expected limiting trajectory of the McKean-Vlasov equation. We therefore find that the two equations indeed have the same expected asymptotic behaviour.

6. Numerical approximations and simulations

In the previous two chapters we investigated the limiting behaviour of the systems we are considering under different limits. In Chapter 4 we investigated what happens when we send the amount of particles N in the interacting particle system to infinity. We found that the trajectory of a fixed particle converges to the trajectory of the McKean-Vlasov equation, when they are coupled with the same Brownian motion. Furthermore, for two particles $i \neq j$, we found that they become asymptotically independent.

In Chapter 5 we considered the limit of $\varepsilon \to 0$. We found that under certain assumptions we can show that the second central moment of the solution of the McKean-Vlasov equation converges to 0 when ε becomes small. Furthermore, we showed that for both systems a LDP holds.

In this chapter we wish to verify these results using numerical approximations to the solutions, as well as using these approximations to study the non-limiting behaviour of the system for specific choices of the functions f and g. For these numerical approximations we will use the Euler-Maruyama scheme, which discretizes the time interval $[0, T]$ into steps of width $h > 0$.

In Section 6.1 we will explain this method further, and show that this approximation is consistent. That is, we show that in the limit $h \to 0$, the approximation converges towards the true solution on a trajectory limit. This implies that for small h the numerical approximations serve as good approximations to the solutions of the original problem.

In Section 6.2 we present various scenarios which we will investigate by doing simulations. These simulations are a simple implementation of the numerical approximations we discus in Section 6.1.

6.1. Euler-Maruyama approximation

The Euler-Maruyama approximation method is a way to approximate solutions of SDEs. The method is similar to that of the Euler (forward) method for ODEs. The method works as follows. Consider the SDE

$$\begin{aligned} \mathrm{d}X(t) &= b(X(t), t)\mathrm{d}t + \sigma(X(t), t)\mathrm{d}B(t), \quad t \in [0, T], \\ X(0) &= x_0. \end{aligned}$$

Now set $n > 0$ and $h = \frac{T}{n}$. Define the timesteps $t_i = hi, i = 0, \dots, n$. The Euler-Maruyama approximation computes the approximations $\hat{X}$ of X at the time steps t_i as follows. First set $\hat{X}(0) = x_0$. Now recursively define

$$\hat{X}(t_i) = \hat{X}(t_{i-1}) + b(\hat{X}(t_{i-1}), t_{i-1})h + \sigma(\hat{X}(t_{i-1}), t_{i-1}) \left(B(t_i) - B(t_{i-1})\right), \quad i = 1, \dots, n.$$

The idea behind this method is that for small h, on the interval $[t_{i-1}, t_i]$, the functions $b(X(t), t)$ do not vary too much, such that they can be approximated by just evaluating them at the left end-point of the interval.

We now wish to show that this approximation indeed works for our model. In order to do this, we first present a couple of different SDEs. We then present how all these equations can be related to each other. We start with the interacting particle model. Let $N \geq 1$ and $\delta, T > 0$ and assume that $\frac{T}{\delta}$ is rational. Consider the SDE

$$
\begin{aligned}
\mathrm{d}X^i(t) &= f(X^i(t))\mathrm{d}t + g\left(\frac{1}{N}\sum_{j=1}^{N} X^j(t-\delta)\right)\mathrm{d}t + \varepsilon \mathrm{d}B^i(t),\\
X^i(0) &= x_0^i,\\
X^i(s) &= \xi_s^i, s \in [-\delta, 0).
\end{aligned}
$$

Also consider the SDE

$$
\begin{aligned}
\mathrm{d}\bar{X}^i(t) &= f(\bar{X}^i(t))\mathrm{d}t + g\left(\mathbb{E}\left[\bar{X}^i(t-\delta)\right]\right)\mathrm{d}t + \varepsilon \mathrm{d}B^i(t),\\
\bar{X}^i(0) &= x_0^i,\\
\mathbb{E}\left[\bar{X}^i(s)\right] &= \xi_s^i, s \in [-\delta, 0).
\end{aligned}
$$

We also consider the local approximations of both SDEs. For $R > 0$, let f_R be such that $f_R = f$ on $B(0, R)$, that f_R satisfies Assumption 1 and that it is globally Lipschitz with Lipschitz coefficient L_R. Consider the equations

$$
\begin{aligned}
\mathrm{d}X_R^i(t) &= f_R(X_R^i(t))\mathrm{d}t + g\left(\frac{1}{N}\sum_{j=1}^{N} X_R^j(t-\delta)\right)\mathrm{d}t + \varepsilon \mathrm{d}B^i(t),\\
X_R^i(0) &= x_0^i,\\
X_R^i(s) &= \xi_s^i, s \in [-\delta, 0),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathrm{d}\bar{X}_R^i(t) &= f_R(\bar{X}_R^i(t))\mathrm{d}t + g\left(\mathbb{E}\left[\bar{X}_R^i(t-\delta)\right]\right)\mathrm{d}t + \varepsilon \mathrm{d}B^i(t),\\
\bar{X}_R^i(0) &= x_0^i,\\
\bar{X}_R^i(s) &= \xi_s^i, s \in [-\delta, 0).
\end{aligned}
$$

Lastly, we consider the Euler-Maruyama approximation for the globally Lipschitz interacting particle model. To do this, let $n \geq 1$, $h = \frac{T}{n}$ such that $h = \frac{\delta}{m}$. Let $t_k = kh, k = -m, \ldots, n$. Consider the approximations.

$$
\begin{aligned}
\hat{X}_R^i(t_k) &= \hat{X}_R^i(t_{k-1}) + f_R(\hat{X}_R^i(t_{k-1}))h + g\left(\frac{1}{N}\sum_{j=1}^{N} \hat{X}_R^j(t_{k-1-m})\right)h\\
&\quad + \varepsilon\left(B^i(t_k) - B^i(t_{k-1})\right), \quad k = 1, \ldots, n,\\
\hat{X}_R^i(0) &= x_0^i,\\
\hat{X}_R^i(t_k) &= \xi_{t_k}^i, \quad k = -m, \ldots, -1.
\end{aligned}
$$

We now also consider an interpolation between these points to make the process continuous. We do this as follows. For $t \in [t_{k-1}, t_k]$, let

$$
\hat{X}_R^i(t) = \hat{X}_R^i(t_{k-1}) + f_R(\hat{X}_R^i t_{k-1})(t - t_{k-1}) + g\left(\frac{1}{N}\sum_{j=1}^{N} \hat{X}_R^j(t_{k-1-m})\right) + \varepsilon(B^i(t) - B^i(t_{k-1})).
$$

We thus use a linear interpolation for the functions f and g, but we use the true trajectory for the Brownian motion. We could also use a linear interpolation for the Brownian motion term, and for small h the difference will be small with high probability, but doing so makes the proof of showing that the approximation is consistent a lot harder. Also set $\hat{X}^i_R(s) = \xi_s$, $s \in [-\delta, 0)$. If we define $j_n(s) = t_n$ for $t_n \leq s \leq t_{n+1}$ then $\hat{X}^i_R$ satisfies the SDE

$$\hat{X}^i_R(t) = x^i_0 + \int_0^t f_R(\hat{X}^i_R(j_n(s))) + g\left(\frac{1}{N}\sum_{j=1}^{N} \hat{X}^j_R(j_n(s-\delta))\right) \mathrm{d}s + \int_0^t \varepsilon \,\mathrm{d}B^i(s),$$

with the appropriate initial conditions.

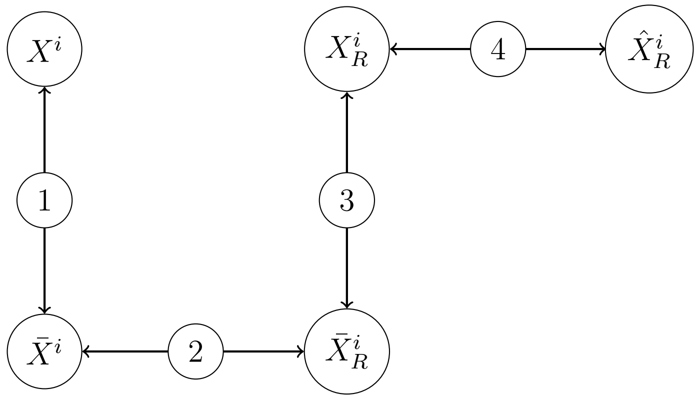

Figure 6.1.: The relation between the different processes.

Let us make clear how all these processes are related. See Figure 6.1. We start with Relation 1. Theorem 4.3 tells us that, when Assumption 1 holds and the initial conditions are i.i.d., there exists a constant M_1, independent of N, such that

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |X^i(t) - \bar{X}^i(t)|^2\right] \leq \frac{M_1}{N}.$$

Next we consider Relation 2. Again assume that Assumption 1 holds. By Theorem 2.19 we have

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |\bar{X}^i(t) - \bar{X}^i_R(t)|^2\right] \leq \mathbb{E}\left[\sup_{0\leq t\leq T} |\bar{X}^i(t) - \bar{X}^i_R(t)|^2 \mathbb{1}\{\tau_R < T\}\right],$$

where τ_R is the first time the process $\bar{X}^i$ leaves the ball $B(0,R)$, which coincides with the exit time of $\bar{X}^i_R$ of the same ball by Theorem 2.19. By Hölder's theorem with $p = q = 2$ we have

$$\leq \left(\mathbb{E}\left[\sup_{0\leq t\leq T} |\bar{X}^i(t) - \bar{X}^i_R(t)|^4\right]\right)^{\frac{1}{2}} \left(\mathbb{P}\left(\tau_R < T\right)\right)^{\frac{1}{2}}.$$

Using the fact that $|a-b|^4 \leq 8|a|^4 + 8|b|^4$ for all $a, b \in \mathbb{R}^d$ we obtain

$$\leq \left(\mathbb{E}\left[\sup_{0\leq t\leq T} 8|\bar{X}^i(t)|^4\right] + \mathbb{E}\left[\sup_{0\leq t\leq T} 8|\bar{X}^i_R(t)|^4\right]\right)^{\frac{1}{2}} \left(\mathbb{P}\left(\tau_R < T\right)\right)^{\frac{1}{2}}.$$

Using Theorem 3.4, stating that the moments of $\bar{X}^i$ and $\bar{X}^i_R$ are finite, and Equation 3.9, with $p = 2$, we have that there exists a constant M_2 independent of R such that

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |\bar{X}^i(t) - \bar{X}^i_R(t)|^2\right] \leq \frac{M_2}{R}.$$

The case of Relation 3 is similar to that of Relation 1. Again assume that Assumption 1 holds and that the initial conditions are i.i.d., we have by using Theorem 4.3 that there exists a constant M_3, independent of N and R, such that

$$\mathbb{E}\left[\sup_{0\le t\le T}|\bar{X}_R^i(t)-X_R^i(t)|^2\right]\le\frac{M_3}{N}.$$

We now only need to investigate Relation 4. We wish to obtain an upper-bound in the form of

$$\mathbb{E}\left[\sup_{0\le t\le T}|X_R^i(t)-\hat{X}_R^i(t)|^2\right]\le M_4 r(h),$$

for some function r which satisfies $\lim_{h\to 0} r(h)=0$. Before we are able to present such a result we first present a lemma describing that in a small time frame, the process X_R^i does not deviate from its initial condition too much. In particular, we have the following result.

Lemma 6.1. *Under the assumptions mentioned above, there exists a constant $\tilde{C}>0$, depending on R, such that for all $t\in[0,T]$ and for all $i=1,\dots,N$ we have*

$$\mathbb{E}\left[\sup_{0\le u\le t}|X_R^i(u)-x_0^i|^2\right]\le\tilde{C}t.$$

Remark. *Instead of looking at how big the deviation from the initial condition is, we can also look at how much a solution at a later time deviates from an earlier time. In particular, let $0\le s<t\le T$, we have*

$$\mathbb{E}\left[\sup_{s\le u\le t}|X_R^i(u)-X_R^i(s)|^2\right]\le\tilde{C}(t-s).$$

This can easily be seen by considering a new SDE, started at time s with initial condition $X_R^i(s)$, and applying the lemma to that process.

Proof. We have that

$$\begin{aligned}\sup_{0\le u\le t}|X_R^i(u)-x_0^i|^2\le 2\sup_{0\le u\le t}&\left|\int_0^u f_R(X_R^i(s))+g\left(\frac{1}{N}\sum_{j=1}^N X_R^j(s-\delta)\right)\,\mathrm{d}s\right|^2\\&+2\sup_{0\le u\le t}\left|\int_0^u\varepsilon\,\mathrm{d}B(s)\right|^2.\end{aligned}$$

Taking expectation, using Theorem 2.13 and Theorem 2.12 yields

$$\begin{aligned}\mathbb{E}\left[\sup_{0\le u\le t}|X_R^i(u)-x_0^i|^2\right]\le 2t\int_0^t\mathbb{E}&\left[2|f_R(X_R^i(s))|^2+2\left|g\left(\frac{1}{N}\sum_{j=1}^N X_R^j(s-\delta)\right)\right|^2\right]\\&+8\int_0^t\varepsilon^2\,\mathrm{d}s.\end{aligned}$$

Using the global Lipschitz assumption of f_R and g, together with the assumption that $f_R(0)=g(0)=0$ to simplify the notation, we obtain

$$\le 2t\int_0^t 2L_R^2\mathbb{E}\left[|X_R^i(s)|^2\right]+2\frac{L^2}{N}\sum_{j=1}^N\mathbb{E}\left[|X_R^j(s-\delta)|^2\right]\,\mathrm{d}s+8\varepsilon^2 t.$$

Using the fact that the second moments of X_R^i are bounded uniformly in time, there exists a constant $C_1 > 0$ such that

$$\begin{aligned}
&\leq 2t \int_0^t C_1 \,\mathrm{d}s + 8\varepsilon^2 t \\
&= C_1 t^2 + 8\varepsilon^2 t \\
&\leq (C_1 T + 8\varepsilon^2) t = \tilde{C} t,
\end{aligned}$$

proving the theorem. ■

We are now ready to present the following theorem.

Theorem 6.2. *Assume that X_R^i and $\hat{X}_R^i$ satisfy the assumptions stated above. Furthermore, assume that for all $s, t \in [-\delta, 0], s < t$ we have that*

$$\mathbb{E}\left[\sup_{s \leq u \leq t} |\xi_u^i - \xi_s^i|^2\right] \leq \tilde{C}(t - s).$$

There exists a constant M_4, depending on R, such that

$$\mathbb{E}\left[\sup_{0 \leq t \leq T} |\hat{X}_R^i(t) - X_R^i(t)|^2\right] \leq M_4 h.$$

Proof. Using Itô's formula we compute

$$\begin{aligned}
|\hat{X}_R^i(t) - X_R^i(t)|^2 = &\int_0^t 2\langle \hat{X}_R^i(s) - X_R^i(s), f_R(\hat{X}_R^i(j_n(s))) - f_R(X_R^i(s))\rangle \,\mathrm{d}s \\
&+ \int_0^t 2\Big\langle \hat{X}_R^i(s) - X_R^i(s)\,, \\
&\quad g\left(\frac{1}{N}\sum_{j=1}^N \hat{X}_R^i(j_n(s-\delta))\right) - g\left(\frac{1}{N}\sum_{j=1}^N X_R^i(s-\delta)\right) \Big\rangle \,\mathrm{d}s.
\end{aligned}$$

Applying Cauchy-Schwarz, the global Lipschitz assumption for both f_R and g and Young's inequality we obtain

$$\begin{aligned}
&\leq \int_0^t 2L_R |\hat{X}_R^i(s) - X_R^i(s)|^2 + L_R |\hat{X}_R^i(j_n(s)) - X_R^i(s)|^2 \,\mathrm{d}s \\
&+ \int_0^t \frac{L_R}{N} \sum_{j=1}^N |\hat{X}_R^j(j_n(s-\delta)) - X_R^j(s)|^2 \,\mathrm{d}s.
\end{aligned}$$

Taking supremum over time and expectation, and using the symmetry of the equations to deduce that X_R^j and $\hat{X}_R^j$ are equally distributed for all j, gives us

$$\begin{aligned}
\mathbb{E}\left[\sup_{0 \leq u \leq t} |\hat{X}_R^i(u) - X_R^i(u)|^2\right] \leq &\int_0^t 2L_R \mathbb{E}\left[|\hat{X}_R^i(s) - X_R^i(s)|^2\right] \,\mathrm{d}s \\
&+ \int_0^t L_R \mathbb{E}\left[|\hat{X}_R^i(j_n(s)) - X_R^i(s)|^2\right] \,\mathrm{d}s \\
&+ \int_0^t L_R \mathbb{E}\left[|\hat{X}_R^i(j_n(s-\delta)) - X_R^i(s-\delta)|^2\right] \,\mathrm{d}s.
\end{aligned}$$

Adding and subtracting the terms $X_R^i(j_n(s))$ and $X_R^i(j_n(s))$ yields

$$\begin{aligned}
&\leq \int_0^t 6L_R\mathbb{E}\left[\sup_{0\leq u\leq s}|\hat{X}_R^i(u)-X_R^i(u)|^2\right]\,\mathrm{d}s \\
&\quad+\int_0^t 2L_R\mathbb{E}\left[|X_R^i(j_n(s))-X_R^i(s)|^2\right]\,\mathrm{d}s \\
&\quad+\int_0^t 2L_R\mathbb{E}\left[|X_R^i(j_n(s-\delta))-X_R^i(s-\delta)|^2\right]\,\mathrm{d}s.
\end{aligned}$$

Applying Grönwall's inequality yields

$$\begin{aligned}
\mathbb{E}\left[\sup_{0\leq t\leq T}|\hat{X}_R^i(t)-X_R^i(t)|^2\right] &\leq e^{6L_RT}\int_0^T 2L_R\mathbb{E}\left[|X_R^i(j_n(s))-X_R^i(s)|^2\right]\,\mathrm{d}s \\
&\quad+e^{6L_RT}\int_0^T 2L_R\mathbb{E}\left[|X_R^i(j_n(s-\delta))-X_R^i(s-\delta)|^2\right]\,\mathrm{d}s.
\end{aligned}$$

We now analyse the first integral. We have by the definition of j_n that

$$\int_0^T \mathbb{E}\left[|X_R^i(j_n(s))-X_R^i(s)|^2\right]\,\mathrm{d}s=\sum_{k=0}^{n-1}\int_{t_k}^{t_{k+1}}\mathbb{E}\left[|X_R^i(t_k)-X_R^i(s)|\right]\,\mathrm{d}s.$$

Using Lemma 6.1 we obtain

$$\begin{aligned}
&\leq\sum_{k=0}^{n-1}\int_{t_k}^{t_{k+1}}\tilde{C}(s-t_k)\,\mathrm{d}s \\
&=\sum_{k=0}^{n-1}\frac{\tilde{C}}{2}h^2 \\
&=n\frac{\tilde{C}}{2}h^2=\frac{\tilde{C}h}{2}.
\end{aligned}$$

By similar arguments we can show that the same holds for the second integral. We therefore have

$$\mathbb{E}\left[\sup_{0\leq t\leq T}|\hat{X}_R^i(t)-X_R^i(t)|^2\right]\leq e^{6L_RT}2\tilde{C}h=M_4h,$$

proving the theorem. ■

We now have shown how the numerical approximations are consistent, in the sense that for each $N\geq 1, R>0$ we have that in the limit $h>0$ the approximation $\hat{X}_R^i$ converges to the process X_R^i on a trajectory level. Furthermore, by choosing R and N large enough, we can make the expected distance between the trajectories of the processes X_R^i, $\bar{X}_R^i$, $\bar{X}^i$ and X^i as small as we desire. We thus have that for sufficiently large R and N and sufficiently small h, the approximation $\hat{X}_R^i$ serves as a good approximation for all of these processes. In the following section we will use this result to analyse certain scenarios using the numerical approximation described above.

6.2. Simulations

Now that we have shown that the numerical approximation is consistent, we can use it to analyse our model in certain scenarios. Throughout this section we will look at specific choices for the functions f and g and the other parameters, and using simulations we will investigate how the model behaves. Throughout these simulations we will be using $d = 1$, as this will help us to better visualise the result of our simulations. In order to simulate the Brownian motions, we make use of the fact that for $0 \leq s < t$, $B(t) - B(s)$ is normally distributed with mean 0 and variance $t - s$. Our simulations are written in `R` (R Core Team (2020)). For the visualisation of our results we make use of the `ggplot2` package (Wickham (2016)) combined with the `scales` package (Wickham and Seidel (2020)). The code used for all of the simulations can be found in Appendix A.

6.2.1. Propagation of chaos

In Chapter 4 we discussed the relation between the interacting particle equation and the McKean-Vlasov equation. In particular, in Theorem 4.3 we found that, when the two processes are coupled, the mean squared distance in the supremum norm converges to 0 when the amount of particles N gets large. We also obtained an upper-bound on this rate of convergence, namely that this decay happens linearly in N. We now wish to analyse numerically how tight this bound is. The expected distance between the two particles of course heavily depends on the choices that we make for the functions f and g, and the parameters such as ε, δ and T. If we, for example, consider the case that $g = c$ for some $c \in \mathbb{R}^d$, then the two equations become the same, such that the distance between the two will be 0 for all values of $N \geq 1$. On the other hand, if the Lipschitz constant of g is large, small differences between the empirical mean of the interacting process and the expectation of the McKean-Vlasov process might still lead to different trajectories, resulting in a larger distance.

In order to test out how this expected squared distance between the trajectories behaves as a function of N, we made the following choices regarding the model. For the functions f and g we will be using

$$f(x) = -x^3 + x, \qquad g(x) = x.$$

It is easy to verify that these functions satisfy Assumption 1 with for example $C = L = 1$ and $L_R = 3R^2 + 1$. The values of the other parameters can be found in Table 6.1.

Parameter	ε	δ	T	h
Value	1	1	10	0.01

Table 6.1.: Values of the parameters.

We will be using a deterministic initial condition. In particular, we will use $X^i(s) = 0, s \in [-\delta, 0], i = 1, \ldots, N$.

If we wish to compare the distance between the trajectory of the interacting particle equation and that of the McKean-Vlasov equation we need to simulate both of the equations. In Section 6.1 we, however, only described a method to simulate the interacting particle equation. Simulation of the McKean-Vlasov equation is difficult, as we have no access to the true expected value of the process. Instead, we will run a simulation of a copy of the interacting particle process Y^{i,N_2} for a high value of N_2, and use this as an estimate for the mean of the McKean-Vlasov process. Theorem 4.3 shows that this method works. The main structure of our simulation is present below (see Algorithm 1). For the amount of particles for the process Y^{i,N_2} we will be using $N_2 = 10.000$. The amount of particles N will be varied, ranging from 1 to 10.000. For each value of N we will repeat the simulation 100 times, to get a good estimate of the mean squared distance. The results of the simulation are presented in Figure 6.2 and

Figure 6.3.

Algorithm 1: Main structure of the simulation

Initialize $Y^i(t_k), k = -m, \dots, 0$
Initialize $\bar{X}^i(t_k), k = -m, \dots, 0$
Initialize $X^i(t_k), k = -m, \dots, 0$
for $k = 1, \dots, n$ **do**
    Compute $Y^i(t_k)$
    Save mean $M_Y(t_k)$
    Compute $\bar{X}^i(t_k)$ using M_Y
    Compute $X^i(t_k)$ using M_X and same noise as $\bar{X}^i(t_k)$
    Save mean M_X
    Save difference $|X^1(t_k) - \bar{X}^1(t_k)|^2$
end
Return $\max_{k=1,\dots,n} |X^1(t_k) - \bar{X}^1(t_k)|$

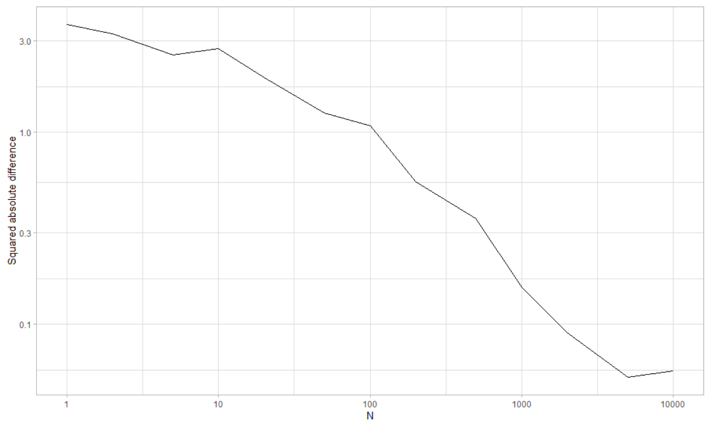

Figure 6.2.: Mean squared distance for different values of N, plotted on a log-log scale.

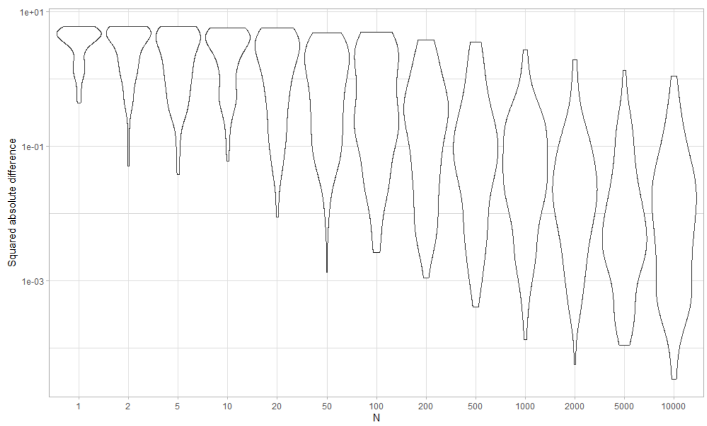

Figure 6.3.: Distribution of squared distances for different values of N.

In Figure 6.2 we plotted the mean squared distance between the trajectories of X^1 and $\bar{X}^1$ in the supremum norm on a log-log plot. In the case that this distance decreases linearly in N, we expect to see a straight line. This is however not the case. It appears that the slope becomes more steep the larger N becomes until N is of the same order as N_2, suggesting an exponential decay rather than a linear one. It can, however, not be justified that this is the case based on this single simulation.

In Figure 6.3 we plotted the distribution of the obtained differences for the different values of N. Notice that the y-axis is on a log-scale. The size of these distributions is normalised such that each one has the same maximum width. Since the sample-size for all values of N is the same, a smaller total area indicated that the distances obtained for that value of N were more concentrated. Based on this graph we can clearly see the decay in distance happening as N increases, not only in the mean but also in the distribution itself. For example, we can also see that the minimum and maximum distance we obtained for each value of N decreases as N increases.

6.2.2. The small noise regime

In Chapter 5 we investigated what happens to the model when ε becomes small. We found that under certain assumptions a LDP holds for both the McKean-Vlasov process as well as the interacting particle process. An important step to show that this result holds is Lemma 5.11, which states that when $\varepsilon \to 0$, we have that the expectation of the McKean-Vlasov process converges to the solution of the associated deterministic equation. We now wish to investigate numerically whether or not a slightly stronger result holds, namely, we investigate the quantity

$$\mathbb{E}\left[\sup_{0\leq t\leq T} |\bar{X}^{\varepsilon}(t) - \varphi(t)|^2\right],$$

with $\bar{X}$ the solution to the McKean-Vlasov equation and φ the solution to the deterministic equation. We will consider the following scenario. Let f and g be defined as

$$f(x) = -x^3 + x, \qquad g(x) = -x.$$

It is easy to verify that f and g satisfy Assumption 1 with $C = L = 1$ and $L_R = 3R^2 + 1$. Therefore we also have that $L_R \leq LR^{\alpha}$, with $L = 4$ and $\alpha = 2$ for $R \geq 1$. The values of the other parameters can be found in Table 6.2.

Parameter	δ	T	h
Value	1	10	0.01

Table 6.2.: Values of the parameters.

The value of ε will vary ranging from 1 to 0.001. We will be using a deterministic initial condition, namely, we set $\bar{X}(s) = \varphi(s) = 1, s \in [-\delta, 0]$.

For the simulation of the McKean-Vlasov process we use a similar approach as we did in Section 6.2.1. We simulate an interacting particle process with $N = 10.000$ and use the mean of this process as an approximation for the mean of the McKean-Vlasov process. This gives our simulation the following structure (see Algorithm 2). For each value of ε, we will repeat the simulation a 100 times to get a good idea of the mean behaviour.

Algorithm 2: Main structure of the simulation

Initialize $Y^i(t_k), k = -m, \ldots, 0$
Initialize $\bar{X}(t_k), k = -m, \ldots, 0$
Initialize $\varphi(t_k), k = -m, \ldots, 0$
for $k = 1, \ldots, n$ **do**
  Compute $Y^i(t_k)$
  Save mean $M_Y(t_k)$
  Compute $\bar{X}(t_k)$ using M_Y
  Compute $\varphi(t_k)$ Save difference $|\bar{X}(t_k) - \varphi(t_k)|^2$
end
Return $\max_{k=1,\ldots,n} |X^1(t_k) - \bar{X}^1(t_k)|$

We first present the trajectory of the solution φ of the deterministic equation in Figure 6.4. The result suggests that the solution of the ODE is that of a (damped) oscillator. We now present the results of our simulations in Figure 6.5 and Figure 6.6.

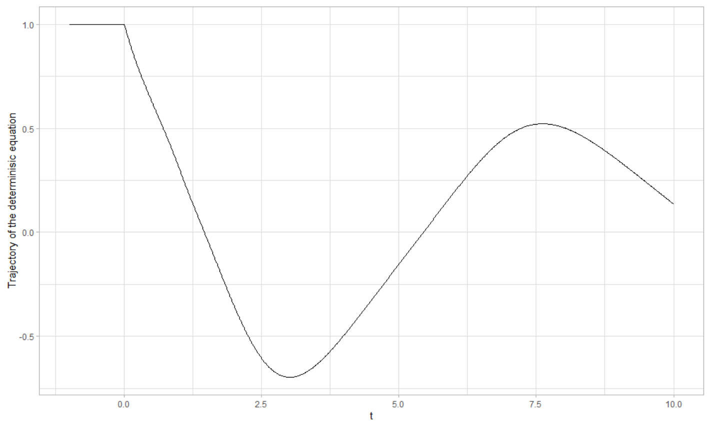

Figure 6.4.: Trajectory of φ, the solution to the ODE.

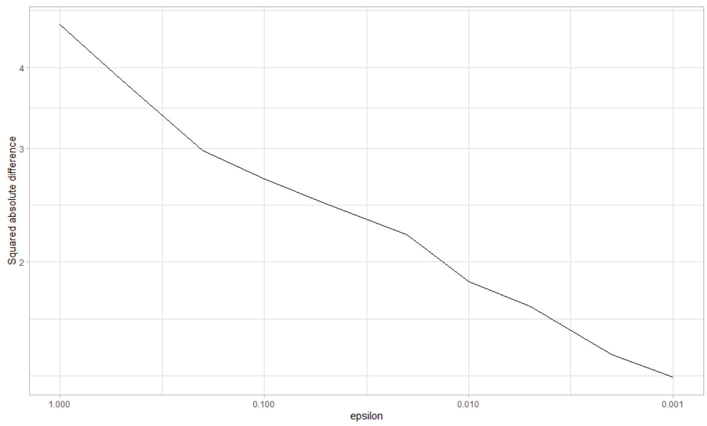

Figure 6.5.: Mean squared distance between the McKean-Vlasov process $\bar{X}$ and φ, the solution to the associated ODE for different values of ε, plotted on a log-log scale.

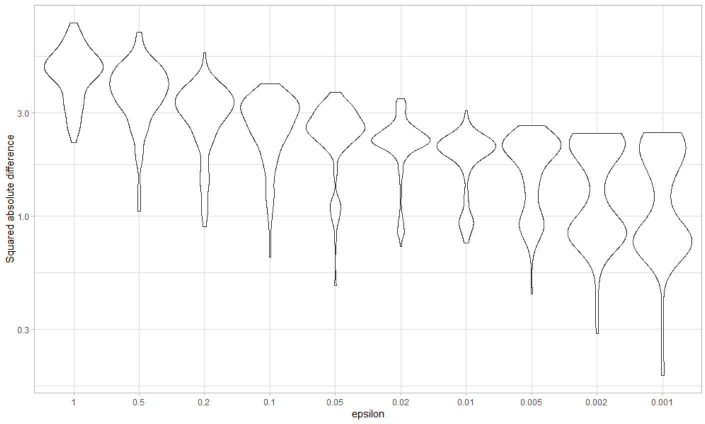

Figure 6.6.: Distribution of the squared distance for various values of ε.

In Figure 6.5 the mean distance between the processes $\bar{X}$ and φ is plotted for various values of ε on a log-log scale. The created line seems linear, indicating that the distance converges to 0 as a power of ε when $\varepsilon \to 0$. This of course does not suffice as a proof, but we conjecture that under certain assumptions a result as above can be proven.

The distributions of the distances obtained for the various values of ε are shown in Figure 6.6. The distributions are normalised such that the maximum width is equal for each value of ε. The y-axis is a log-scale. The convergence that we saw for the mean can also be seen in the distributions. For example, the maximum and minimum distances that we obtained per ε decrease as ε becomes smaller.

6.2.3. The impact of delay

In this section we will show that having an interaction with delay fundamentally changes the model. We will be showing this for the ODE

$$\frac{\mathrm{d}\varphi(t)}{\mathrm{d}t} = f(\varphi(t)) + g(\varphi(t-\delta)), \quad t \in [0,T],$$
$$\varphi(s) = x_0, \quad s \in [-\delta, 0].$$

For $a > 0$, define the function $h_a : \mathbb{R} \to \mathbb{R}$ as

$$h_a(x) = \begin{cases} -1, & x < -a, \\ \frac{x}{a}, & -a \leq x \leq a, \\ 1, & a < x. \end{cases}$$

Put into words, h_a is the sign function but smoothed in the interval $[-a, a]$ in order to make sure that it satisfies a Lipschitz assumption. We now consider two scenarios; in the first scenario we set $f = h_a$ and $g = 0$, such that the system only depends on the current state without influences from the past. We test two values of a, namely $a = 0.5$ and $a = 1$. The values of the other parameters can be found in Table 6.3.

Parameter	δ	T	x_0
Value	1	25	0.1

Table 6.3.: Values of the parameters.

The resulting trajectory for φ can be seen in Figure 6.7 and Figure 6.8 for $a = 0.5$ and $a = 1$ respectively. We clearly have that $\varphi(t) \to 0$ as $t \to \infty$. As long as $0 \leq x_0 \leq a$, we can even find an explicit expression for φ. It can be shown easily that in this case φ is given by

$$\varphi(t) = x_0 e^{-\frac{t}{a}}, \quad t \geq 0.$$

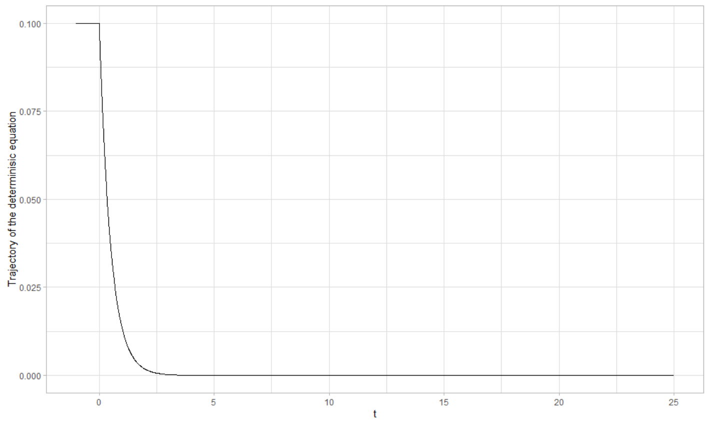

Figure 6.7.: Trajectory of φ with $f = h_{0.5}$ and $g = 0$.

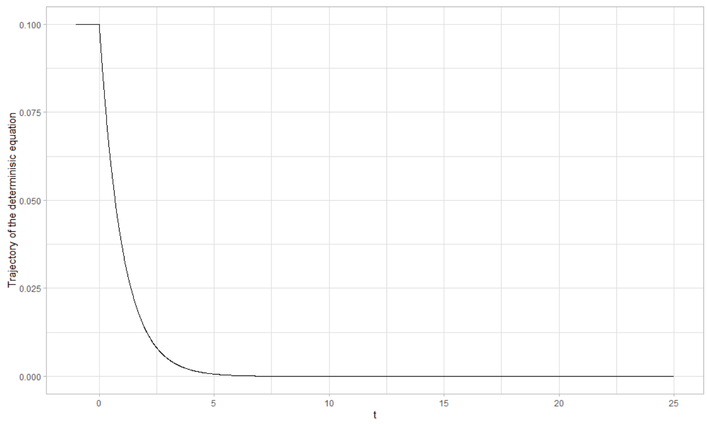

Figure 6.8.: Trajectory of φ with $f = h_1$ and $g = 0$.

Let us now switch the role of the functions f and g. That is, we now set $f = 0$ and $g = h_a$. We keep the other parameters the same, and test the same values of a. The results can be seen in Figure 6.9 and 6.10 for $a = 0.5$ and $a = 1$ respectively. We now see that the two different values of a lead to different limiting behaviours. For $a = 0.5$, the system converges towards a periodic function, with - in this case - an amplitude larger than the initial condition. On the other hand, for $a = 1$, the system converges towards 0. In contrary to the scenario without delay, this convergence is no longer monotonic. The critical value a^*, the value for which the system becomes periodic when $a < a^*$ and converges towards 0 when $a > a^*$ seems to be around 0.64. The exact value is however hard to determine, and likely also depends on the parameters used, such as δ and x_0.

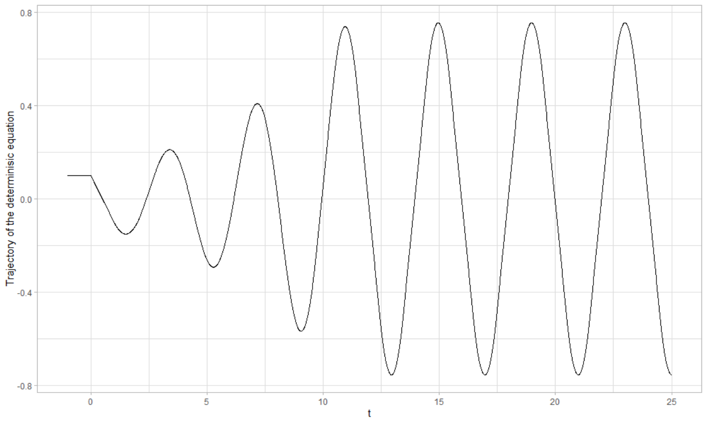

Figure 6.9.: Trajectory of φ with $f = 0$ and $g = h_{0.5}$.

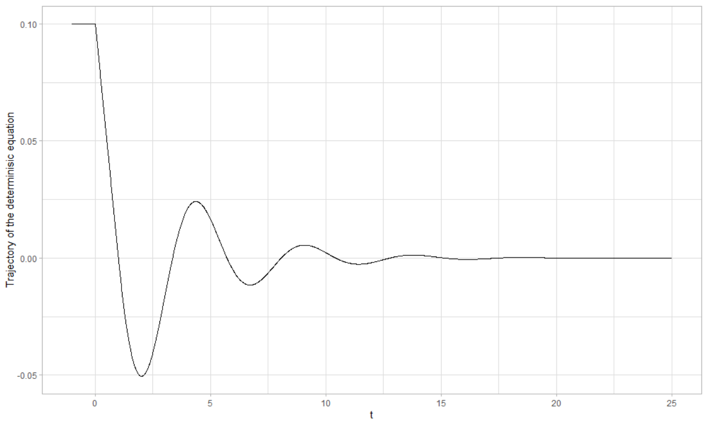

Figure 6.10.: Trajectory of φ with $f = 0$ and $g = h_1$.

During our research for this thesis we have investigated many more scenarios, as mentioned in the introduction. In particular, we tried to find criteria for when the mean of either the interacting particle process or the McKean-Vlasov process becomes a periodic function in time. We are however unable to formulate a clear answer to this question. We are also not certain which parts of the results we obtained are due to the models themselves and which parts are mere artifacts of our numeric approach used. Therefore, we decided to not include these results in this thesis.

7. Conclusion

7.1. Summary

Throughout this thesis we investigated a SDE describing an interacting particle process where the interaction happens with a delay through the empirical mean, and a delayed McKean-Vlasov SDE. We formulated assumptions under which these SDEs posses unique solutions. In particular, we showed that for the drift function f a one-sided Lipschitz assumption combined with a local Lipschitz function suffices. This assumption is more general than the sub-linear growth assumption often made. We also showed that all the moments of both processes in the supremum norm are finite, under the additional assumptions that g is bounded for the interacting particle process.

Under the same assumptions as we made for the existence, and the assumption that the initial values for the interacting particle process are i.i.d., we show that the propagation of chaos property holds. That is, the trajectories of the interacting particle model converge to those of the McKean-Vlasov equation under a suitable coupling when the amount of particles N is send to infinity. This also implies that the trajectories of two particles $i \neq j$ become asymptotically independent in the same limit.

In Chapter 5 we provided assumptions for both models under which a LDP hold. We also derive an explicit expression for the rate functions for both models. Although these rate functions are different for both models, we show that the expected asymptotic behaviour is the same for both models in the limit $\varepsilon \to 0$.

Lastly, we show that the Euler-Maruyama approximation scheme is consistent for the interacting particle process when we assume that the function f is globally Lipschitz. That is, in the limit of sending the step size h used in the approximation to 0, the approximation converges towards the true process in the supremum norm. We use this result to study a series of scenarios using simulations. We investigate how some theoretical bounds obtained in this thesis compare to real scenarios. We also briefly discuss how the delay term can lead to a periodic function, while the model would converge to a constant without the delay.

7.2. Discussion

In this thesis we investigated multiple aspects of the equations we have been studying. By focusing more on a specific aspect it is likely to obtain more general results than we have obtained. There are a variety of methods to generalise our results. Firstly, we only considered the case with additive noise. By multiplying the Brownian motion with a function σ, depending on the process, possibly even with a delay, a more general noise term can be introduced. We suspect that under certain assumptions on this function σ, similar results as we have described in this thesis could be obtained. Secondly, the functions f and g in our model do not explicitly depend on time. Results for standard SDEs, as described in Chapter 2, suggest that introducing this explicit dependence does not require many extra assumptions. We thus expect that our results can still be shown for functions explicitly depending on time without making too many extra assumptions. Finally, it might be possible to relax the assumptions we made in this thesis. In particular, the global Lipschitz assumption on the function g is rather strict, and we suspect that this assumption can be weakened for most of the results we have presented in this thesis.

Another interesting take for further research would be to focus more on the non-limiting behaviour of the model. During our research we investigated many different scenarios, but we were unable to formulate clear conclusions. Particular interesting questions are in which scenario the system becomes periodic, i.e. the mean of the process is a periodic function in time and in which scenarios the mean converges towards one fixed point. In the case that the mean is a periodic function over time, questions such as how the period and the amplitude depend on the parameters used could be of interest. In the case that the mean converges to a fixed point, the rate of convergence could be investigated. In particular, whether this rate is exponential or not.

Bibliography

Paolo Baldi. *Stochastic calculus*. Springer, 2017.

Paolo Baldi and Lucia Caramellino. General freidlin–wentzell large deviations and positive diffusions. *Statistics & Probability Letters*, 81(8):1218–1229, 2011.

Alberto Chiarini and Markus Fischer. On large deviations for small noise itô processes. *Advances in Applied Probability*, 46(4):1126–1147, 2014.

Iosif Ilyich Gikhman and Anatoliy Volodymyrovych Skorohod. *Stochastic Differential Equations*. Springer, 1972.

Samuel Herrmann, Peter Imkeller, Ilya Pavlyukevich, and Dierk Peithmann. *Stochastic Resonance: A Mathematical Approach in the Small Noise Limit*. American Mathematical Society, 2014.

Xuerong Mao. *Stochastic differential equations and applications*. Elsevier, 2007.

Henry P. McKean. A class of markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences of the United States of America*, 56(6):1907, 1966.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020. URL `https://www.R-project.org/`.

Markus Reiß. Stochastic differential equations, February 2007.

Alain-Sol Sznitman. Topics in propagation of chaos. In Paul-Louis Hennequin, editor, *Ecole d'Eté de Probabilités de Saint-Flour XIX — 1989*, pages 165–251, Berlin, Heidelberg, 1991. Springer Berlin Heidelberg.

Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4. URL `https://ggplot2.tidyverse.org`.

Hadley Wickham and Dana Seidel. *scales: Scale Functions for Visualization*, 2020. URL `https://CRAN.R-project.org/package=scales`. R package version 1.1.1.

A. Code used for simulations

```r
# Package required to create the plots
library(ggplot2)
library("scales")
reverselog_trans <- function(base = exp(1)) {
  trans <- function(x) -log(x, base)
  inv <- function(x) base^(-x)
  trans_new(paste0("reverselog-", format(base)), trans, inv,
            log_breaks(base = base),
            domain = c(1e-100, Inf))
}

# Main simulation
Simulation <- function(x0,epsilon,Tmax,delta,stepcount,N,NMcKean){
  # Initialize all the data
  h <- delta/stepcount
  tVec <- seq(-delta,Tmax,h)
  indepVec <- rep(x0,NMcKean)
  xVec <- rep(x0,N)
  x2Vec <- rep(x0,N)
  McKeanVec <- rep(0,length(tVec))
  McKeanVec[1:(stepcount+1)] <- x0
  MeanVec <- McKeanVec
  MeanindepVec <- McKeanVec
  MeanMcKeanVec <- McKeanVec
  DeterVec <- McKeanVec
  Diff2 <- rep(0,length(tVec))
  Diff2[1:(stepcount+1)] <- 0
  pb <- winProgressBar(title = "progress bar", min = 1,
                       max = length(MeanVec), width = 300)
  # Compute all the timesteps
  for(i in (stepcount+2):length(MeanVec)){
    indepVec <- indepVec +  h*f(indepVec) + h*g(MeanindepVec[i-stepcount-1]) + sqrt(epsilon*h)*rnorm(NMcKean,0,1)
    MeanindepVec[i] <- mean(indepVec)
    noise <- rnorm(N,0,1)
    McKeanVec[i] <- McKeanVec[i-1] + h*f(McKeanVec[i-1]) + h*g(MeanindepVec[i-stepcount-1]) + sqrt(epsilon*h)*noise[1]
    x2Vec <- x2Vec +  h*f(x2Vec) + h*g(DeterVec[i-stepcount-1]) + sqrt(epsilon*h)*noise
    MeanMcKeanVec[i] <- mean(x2Vec)
    xVec <- xVec +  h*f(xVec) + h*g(MeanVec[i-stepcount-1]) + sqrt(epsilon*h)*noise
    MeanVec[i] <- mean(xVec)
    Diff2[i] <- (McKeanVec[i]-xVec[1])^2
    DeterVec[i] <- DeterVec[i-1] + h*f(DeterVec[i-1]) + h*g(DeterVec[i-stepcount-1])
    setWinProgressBar(pb,i,title = paste(round(i/length(MeanVec)*100,1),"% done"))
  }
  # Close the progress bar and return the results
```

```r
  close(pb)
  return(list(tVec,MeanVec,McKeanVec,Diff2,DeterVec,MeanMcKeanVec))
}

## Propagation of chaos
# Define the functions f and g
f <- function(x){
  -x^3+x
}

g <- function(x){
  x
}
# Specify the parameter used
x0 <- 0
epsilon <- 1
delta <- 1
Tmax <- 10*delta
stepcount <- 100
NVec <- c(1,2,5,10,20,50,100,200,500,1000,2000,5000,10000)
NMcKean <- 10000
runs <- 100

# Perform the simulation
MaxMatrix <- matrix(nrow = runs,ncol = length(NVec))
for(j in 1:length(NVec)){
  for(i in 1:runs){
    result <- Simulation(x0,epsilon,Tmax,delta,stepcount,NVec[j],NMcKean)
    MaxMatrix[i,j] <- max(result[[4]])
  }
  print(paste(runs,"runs done for N =",NVec[j]))
}

# Convert the data
Ncol <- rep(NVec,each=runs)
DiffCol <- c()
for(i in 1:length(NVec)){
  DiffCol <- c(DiffCol,MaxMatrix[,i])
}
MeanCol <- rep(colMeans(MaxMatrix),each=runs)
MyData <- data.frame(Ncol,DiffCol,MeanCol)

# Plot the data
ggplot(MyData, aes(Ncol,MeanCol)) + geom_line()  +
  scale_x_log10() + scale_y_log10() +
  theme_light() + labs(x="N",y="Squared absolute difference")

ggplot(MyData,aes(as.factor(Ncol),DiffCol)) + geom_violin(scale="width") + theme_light()
    +
  labs(x="N",y="Squared absolute difference") + scale_y_log10()
```

```r

## Small noise regime
# Define the functions f and g
f <- function(x){
  -x^3 + x
}

g <- function(x){
  -x
}

# Specify the parameters
x0 <- 1
delta <- 1
Tmax <- 10*delta
stepcount <- 100
NMcKean <- 10000
N <- 10000
epsilonVec <- c(1,0.5,0.2,0.1,0.05,0.02,0.01,0.005,0.002,0.001)
runs <- 100

# Perform the simulation
DistMatrix <- matrix(nrow = runs,ncol = length(epsilonVec))
for(j in 1:length(epsilonVec)){
  for(i in 1:runs){
    result <- Simulation(x0,epsilonVec[j],Tmax,delta,stepcount,1,NMcKean)
    DistMatrix[i,j] <- max((result[[3]]-result[[5]])^2)
  }
  print(paste(runs,"runs done for epsilon =",NVec[j]))
}

# Convert the data
epsiloncol <- rep(epsilonVec,each=runs)
DiffCol <- c()
for(i in 1:length(epsilonVec)){
  DiffCol <- c(DiffCol,DistMatrix[,i])
}
MeanCol <- rep(colMeans(DistMatrix),each=runs)
MyData <- data.frame(epsiloncol,DiffCol,MeanCol)

# Plot the data
ggplot(MyData, aes(epsiloncol,MeanCol)) + geom_line()  + scale_x_continuous(trans=reverselog_trans(10)) +
  scale_y_log10() + theme_light() + labs(x="epsilon",y="Squared absolute difference")

ggplot(MyData,aes(factor(epsiloncol,levels=epsilonVec),DiffCol)) + geom_violin(scale="width") +
  scale_y_log10() + theme_light() + labs(x="epsilon",y="Squared absolute difference")

## The impact of delay
```

```r
# Define the functions f and g
f <- function(x){
  0
}

g <- function(x){
  -( sign(x)*(x< -a) + x/a*(x>=-a)*(x<=a) + sign(x)*(x>a) )
}

# Specify the parameters
a <- 1
x0 <- 0.1
delta <- 1
Tmax <- 25*delta
stepcount <- 100
NMcKean <- 1
N <- 1
epsilon <- 0

# Perform the simulation
result <- Simulation(x0,epsilon,Tmax,delta,stepcount,N,N)

# Convert the data
MyData <- data.frame(time=result[[1]],McKean=result[[3]],Deter=result[[5]],Mean=result[[6]],Interacting=result[[2]])

# Plot the data
ggplot(MyData, aes(time,Deter)) + geom_line() + theme_light() + labs(x="t",y="Trajectory of the determinisic equation")
```