On the uniqueness of a solution and stability of McKean-Vlasov stochastic differential equations

Jani Nykänen

Master's thesis
University of Jyväskylä
Department of mathematics and statistics

January 19, 2020

Tiivistelmä

Jani Nykänen, *On the uniqueness of a solution and stability of McKean-Vlasov stochastic differential equations*, Jyväskylän yliopisto, Matematiikan ja tilastotieteen laitos, matematiikan pro gradu -tutkielma, 75 s., tammikuu 2020.

Tässä tutkielmassa tutustutaan McKeanin-Vlasovin stokastisiin differentiaaliyhtälöihin, jotka yleistävät tavalliset stokastiset differentiaaliyhtälöt lisäämällä kerroinfunktioihin riippuvuuden tuntemattoman prosessin jakaumasta tietyllä ajanhetkellä. Pääasiallisena lähteenä seurataan K. Bahlalin, M. Mezerdin ja B. Mezerdin artikkelia *Stability of Mckean-Vlasov stochastic differential equations and applications*.

Tutkielmassa käydään läpi tarvittavia esitietoja todennäköisyysteoriasta ja tavallisista stokastisista differentiaaliyhtälöistä. Kerroinfunktioiden jatkuvuuden ja mitallisuuden määrittämiseksi esitellään Wassersteinin etäisyys, joka on metriikka äärellismomenttisten reaaliavaruuden todennäköisyysmittojen avaruudessa. Metriikan avulla saadaan yleistettyä lause, joka takaa ratkaisun olemassaolon ja yksikäsitteisyyden, kun kerroinfunktiot ovat Lipschitz-jatkuvia ja toteuttavat lineaarisen kasvuehdon. Lisäksi osoitetaan, että yksikäsitteisyys on voimassa eräällä Lipschitz-jatkuvuutta heikommalla ehdolla.

Numeerisessa ratkaisemisessa voidaan hyödyntää tulosta, jossa konstruoidaan iteroitu jono prosesseja, jotka suppenevat kohti yksikäsitteistä ratkaisua. Lopuksi tarkastellaan ratkaisuprosessien stabiiliutta erikseen alkuarvon, kerroinfunktioiden ja integroivan prosessin suhteen.

Abstract

In this thesis we introduce McKean-Vlasov stochastic differential equations, which are a generalization of ordinary stochastic differential equations, but now the coefficients depend on the distribution of the unknown process. In our main results we follow K. Bahlali, M. Mezerdi and B. Mezerdi's article *Stability of Mckean-Vlasov stochastic differential equations and applications.*

We start by giving preliminary theory required to understand our main results. To define continuity and measurability of the coefficient functions, we introduce the Wasserstein distance, which is a metric in the space of probability measures on the real line with finite moments. With the metric we generalize a theorem that states that a unique solution exists provided that the coefficients are Lipschitz continuous and satisfy the linear growth condition. In addition we show that in a specific case the uniqueness holds even if the coefficients satisfy a condition weaker than Lipschitz continuity.

In numerics one can use a result that provides a way to approximate the solution with a sequence of iterated processes converging to the unique solution. In the last part we consider stability of the solution with respect to the initial value, the coefficients and the driving process.

Contents

1 Introduction

In this thesis we consider a class of stochastic differential equations, where the coefficient functions depend on the law of the solution process. This class is called McKean-Vlasov stochastic differential equations (MVSDE). Compared to classical stochastic differential equations, the distribution variable adds another layer of complexity. For instance, many tools from classical stochastic calculus cannot be directly applied to study properties of these equations or computing solutions.

When modeling real-life phenomena with mathematical models, one notable problem with ordinary differential equations is the lack of randomness, which sometimes leads to situations where the given data cannot be matched with a function that is obtained as the unique solution to a differential equation. In equations that model particle systems this was a particular problem. To overcome this issue probability and differential equations were joined together [Sob91, introduction].

In its modern form the theory of stochastic differential equations — and stochastic calculus in general — was created by Kiyosi Itô. His first paper in stochastic integration was published in 1944. Itô defined stochastic integrals with respect to the Brownian motion, but his definition was generalized to semimartingales by J. L. Doob in 1953. In 1951 Itô published an influential paper, where he stated and proved a formula, later known as Itô's formula (see section 2.4.2), one of the most powerful tools in stochastic calculus [Mey09], [JP04].

Like ordinary stochastic differential equations, the origins of McKean-Vlasov stochastic differential equations are in physics. The equations were studied to obtain a model for a large system of weakly interacting particles, when the number of particles tends to infinity. In a way, this gives the average behaviour of one particle [BMM19].

The original Vlasov equation modeled the interactions of a system of particles in plasma [PCMM15]. In 1956 M. Kac published a paper where he studied a stochastic counterpart of Vlasov's equation in the context of statistical physics [Kac56], [BMM19]. A probabilistic formulation for this equation was given by H. P. McKean in 1966, when he considered the problem from the perspective of Markov processes. He formulated the problem as a stochastic differential equation, where the coefficients depended upon the expected value of the unknown process [McK66].

Recently MVSDEs have gained attention in the theory of mean-field games, which is a branch of game theory. Mean-field games model strategic decision games with a large population of players, usually called agents, who try to choose an optimal strategy when they only have macroscopic information of the game, resulted by the other players. MVSDEs can be used to model this situation as the number of players tends to infinity. This theory generalizes the applications of MVSDEs from physics to economics and finance, see for instance [LL07].

Our primary goal in this thesis is to generalize the theory, known for ordinary stochastic differential equations, to the setting of McKean-Vlasov SDEs by following [BMM19]. We formulate and prove most theorems of the article, but in more detail and in some cases with different assumptions. In addition we contribute by demonstrating every result with at least one example.

We start by recalling some preliminary theory to understand our main results in Section 2. Then we introduce the Wasserstein distance, which is a metric in the space of probability measures. It allows us to generalize existence and uniqueness theorems for MVSDEs. Then we focus on approximation and stability theorems. We prove an iterative method for approximating and computing the unique solution of an MVSDE. We consider three different stability results. In the first case we show that a map between initial values and their corresponding solutions is continuous. In the second stability result we approximate the solution by approximating the coefficients. The last stability theorem states that we can approximate the solution also by approximating the driving process with continuous martingales.

2 Preliminaries

We start by introducing some definitions and theorems that are used within this thesis. The reader is assumed to be familiar with common topological concepts and measure theory, but not necessarily probability theory. We give the background for ordinary stochastic differential equations, which serves as a basis for our main results in this thesis.

Throughout this section, if we assume an index set $I \subset \mathbb{R}$, we assume that $I = [0, T]$ for some $T > 0$.

2.1 Notation and terminology

Here we list some essential terminology and notation used throughout this thesis.

- A function $f : \mathbb{R} \to \mathbb{R}$ is called *increasing*, if $s < t$ implies $f(s) \leq f(t)$ for all $s, t \in \mathbb{R}$. If one has $f(s) < f(t)$ for all $s < t$, then f is called *strictly increasing*. Respectively, if $f(s) \geq f(t)$, then the function f is called *decreasing*, and if $f(s) > f(t)$, *strictly decreasing*.

- By natural numbers we mean $\mathbb{N} = \{1, 2, 3, ...\}$. In particular, $0 \notin \mathbb{N}$.
- If X and Y are topological spaces, we denote by $\mathcal{C}(X, Y)$ the space of continuous maps from X to Y.
- The indicator function over a set A is defined by

$$\mathbb{1}_A(x) := \begin{cases} 1, & x \in A \\ 0, & x \notin A. \end{cases}$$

- The power set of a non-empty set X is

$$2^X := \{B \mid B \subseteq X\}.$$

  In particular $\emptyset, X \in 2^X$.
- We denote by $\lfloor \cdot \rfloor$ the floor function, that is

$$\lfloor x \rfloor := \max \{z \in \mathbb{Z} \mid x \geq z\}$$

  for $x \in \mathbb{R}$.
- By $\mathbb{R}^{m \times n}$ we denote $m \times n$-matrices of which components are real-valued. For $A = [a_{i,j}] \in \mathbb{R}^{m \times n}$ we use the matrix norm

$$\|A\| = \sqrt{\sum_{\substack{i=1,...,m \\ j=1,...,n}} |a_{i,j}|^2}.$$

- In general, if not stated otherwise, we use $\|\cdot\|$ to denote the Euclidean norm, that is,

$$\|x\| = \|(x_1, ..., x_d)\| = \sqrt{\sum_{k=1}^{d} |x_k|^2}$$

  for $x = (x_1, ..., x_d) \in \mathbb{R}^d$.
- We denote by $\langle \cdot \mid \cdot \rangle$ the inner product in $\mathbb{R}^d$, that is,

$$\langle x \mid y \rangle = \sum_{k=1}^{d} x_k y_k$$

  for $x, y \in \mathbb{R}^d$.

2.2 Probability theory

In this section we introduce the measure theoretical basis of modern probability theory.

2.2.1 Stochastic basis

We start by recalling definitions for measure spaces and stochastic bases. (See, for example, [GG18, definitions 1.1.1 and 1.3.2])

Let Ω be a non-empty set and $\mathcal{F} \subseteq 2^{\Omega}$. The set $\mathcal{F}$ is a σ-algebra, if it satisfies the following conditions:

(1) $\emptyset, \Omega \in \mathcal{F}$

(2) If $A \in \mathcal{F}$, then $A^{\complement} = \Omega \setminus A \in \mathcal{F}$.

(3) Let $A_1, A_2, A_3, ... \in \mathcal{F}$. Then

$$\bigcup_{k=1}^{\infty} A_k \in \mathcal{F}.$$

The pair $(\Omega, \mathcal{F})$ is called *measurable space*. If $A \in \mathcal{F}$, the set A is called *measurable*. A map $\mu : \Omega \to \mathbb{R}$ is a *measure* on $(\Omega, \mathcal{F})$, if the following holds:

(1) $\mu(\emptyset) = 0$.

(2) If $A_1, A_2, ... \in \mathcal{F}$ be pair-wise disjoint sets, that is, $A_k \cap A_j = \emptyset$ for all $k \neq j$, then

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k).$$

The triplet $(\Omega, \mathcal{F}, \mu)$ is called *measure space*. Moreover, if $\mu(\Omega) = 1$, the measure μ is called *probability measure*, and the space $(\Omega, \mathcal{F}, \mu)$ *probability space*.

Definition 2.1 (Filtration, [Gei19, definition 2.1.8])**.** Assume a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $I \subseteq \mathbb{R}$ be an index set and $(\mathcal{F}_t)_{t\in I}$ a family of sub-σ-algebras of $\mathcal{F}$, that is, for all $t \in I$, $\mathcal{F}_t$ is a σ-algebra and $\mathcal{F}_t \subseteq \mathcal{F}$. The family $(\mathcal{F}_t)_{t\in I}$ is a *filtration*, if it satisfies

$$\mathcal{F}_s \subseteq \mathcal{F}_t$$

for all $s, t \in I$ with $s \leq t$. The quadruple $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in I})$ is called *stochastic basis*.

Remark 2.2. As mentioned earlier, we assume that the index set is the interval $[0, T]$ for some $T > 0$, but in general it could be, for instance, $\mathbb{N}$ or a subset.

In applications, it is usually required that the stochastic basis satisfies some specific properties known as *the usual conditions*. Before we can define these conditions, we must recall the definition of a complete measure space.

Definition 2.3 ([GG18, definition 1.6.1])**.** A measure space $(\Omega, \mathcal{F}, \mu)$ is *complete*, if every subset of every null set is measurable, that is, if $N \in \mathcal{F}$ and $\mu(N) = 0$, then for all $S \subseteq N$ it holds that $S \in \mathcal{F}$.

Definition 2.4 (Usual conditions, [Gei19, definition 2.4.11])**.** The stochastic basis $(\Omega, \mathcal{F}, \mathbb{P}, (F_t)_{t\in I})$ satisfies *the usual conditions*, if the following holds:

(1) $(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space.

(2) If $A \in \mathcal{F}$ and $\mathbb{P}(A) = 0$, then $A \in \mathcal{F}_t$ for all $t \in I$.

(3) The filtration is right-continuous, that is,

$$\mathcal{F}_t = \bigcap_{\substack{\epsilon>0 \\ t+\epsilon\in I}} F_{t+\epsilon}$$

for all $t \in [0, T)$

2.2.2 Borel σ-algebra and Lebesgue measure

Assume a topological space X. The *Borel σ-algebra on E*, denoted by $\mathcal{B}(X)$ is the smallest σ-algebra containing all open sets on X. It follows by the definition of σ-algebra that $\mathcal{B}(X)$ also contains every closed set of X. A construction for the Borel σ-algebra can be found in [Rud70, chapter 1]. In this thesis we consider only the case where X is a separable Banach space.

If μ is a measure on $(X, \mathcal{B}(X))$, then it is called *Borel measure*. An important example of a Borel measure on $\mathbb{R}$ is the Lebesgue measure λ. For any half-open interval one has

$$\lambda((a, b]) = b - a,$$

if $b > a$. More generally, if λ^d is a d-dimensional Lebesgue measure, it gives the geometric measure of any Borel set $A \subseteq \mathbb{R}^d$. A construction for the Lebesgue measure is given in [Rud70, chapter 2, theorem 2.20].

2.2.3 Random variables

Assume two measurable spaces $(\Omega, \mathcal{F})$ and $(\Gamma, \mathcal{G})$. The map $f : \Omega \to \Gamma$ is $(\mathcal{F}, \mathcal{G})$-*measurable*, if

$$f^{-1}(B) \in \mathcal{F}$$

for all $B \in \mathcal{G}$. If X is a topological space and $(\Gamma, \mathcal{G}) = (X, \mathcal{B}(X))$, then we call $(\mathcal{F}, \mathcal{B}(X))$-measurable maps $\mathcal{F}$-measurable or just measurable, if the σ-algebra $\mathcal{F}$ is clear from the context.

If X and Y are topological spaces, then we call a map $f : X \to Y$ *Borel measurable*, if it is $(\mathcal{B}(X), \mathcal{B}(Y))$-measurable.

If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and E is a separable Banach space, measurable maps $g : \Omega \to E$ are called *random variables*. The *law of a random variable* g is the measure $\mathbb{P}_g$ on $(E, \mathcal{B}(E))$ defined by

$$\mathbb{P}_g(A) := \mathbb{P}(g \in A) = \mathbb{P}(\{\omega \in \Omega \mid g(\omega) \in A\})$$

for all $A \in \mathcal{B}(E)$. We notice that $\mathbb{P}_g(E) = \mathbb{P}(g \in E) = 1$, thus $\mathbb{P}_g$ is a probability measure.

2.2.4 Integration theorems

Here we introduce intergration theorems we need in our proofs.

Assume a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $f : \Omega \to \mathbb{R}$ be a random variable. We denote the *expected value of a random variable* f by

$$\mathbb{E}f = \int_\Omega f(\omega) \, \mathrm{d}\mathbb{P}(\omega),$$

where the right-hand side denotes the integral with respect to the measure $\mathbb{P}$, assuming that it is finite. For more information and properties, see [GG18, definition 5.1.3-5.1.4].

If it is clear from the context, we may use the shorter notation

$$\int_\Omega f(\omega) \, \mathrm{d}\mathbb{P}(\omega) = \int_\Omega f \, \mathrm{d}\mathbb{P}.$$

If $\mathbb{P}_f$ is the law of the random variable f, we have

$$\int_\Omega f(\omega) \, \mathrm{d}\mathbb{P}(\omega) = \int_\mathbb{R} x \, \mathrm{d}\mathbb{P}_f(x).$$

For an $\mathbb{R}^d$-valued random variable $g : \Omega \to \mathbb{R}^d$, $g(\omega) = (g_1(\omega), ..., g_d(\omega)) \in \mathbb{R}^d$, we define

$$\mathbb{E}g = \left(\int_\Omega g_1 \, \mathrm{d}\mathbb{P}, ..., \int_\Omega g_d \, \mathrm{d}\mathbb{P} \right) \in \mathbb{R}^d,$$

assuming that the integrals exist.

For all $p > 0$ we say that $f \in \mathcal{L}_p(\Omega, \mathcal{F}, \mathbb{P})$, if

$$\mathbb{E}\, \|f\|^p < \infty,$$

assuming that $f : \Omega \to \mathbb{R}^d$ is a random variable.

If λ denotes the Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then we use the notation

$$\int_{[a,b]} g(x)\, \mathrm{d}\lambda(x) = \int_{\mathbb{R}^d} \mathbb{1}_{[a,b]}(x) g(x)\, \mathrm{d}\lambda(x) = \int_a^b g(x)\, \mathrm{d}x$$

for an integrable function $g : [a, b] \to \mathbb{R}$. This notation is usually reserved for the Riemann integral. However, under certain conditions the Riemann integral and the integral with respect to the Lebesgue measure coincides.

Proposition 2.5 ([GG18, Proposition 5.5.1])**.** *Let $g : [a, b] \to \mathbb{R}$ be a bounded and Borel measurable function. Assume that there exists a set $\mathcal{N} \in \mathcal{B}(\mathbb{R})$ with $\mathcal{N} \subseteq [a, b]$ and $\lambda(\mathcal{N}) = 0$ such that g is continuous in $[a, b] \setminus \mathcal{N}$. Then g is Riemann-integrable and*

$$\int_a^b g(x)\, \mathrm{d}x = \int_{[a,b]} g(x)\, \mathrm{d}\lambda(x),$$

where the left-hand side denotes the Riemann integral.

This theorem justifies the notation. In the cases where we use the Riemann integral, we mention it separately.

We continue with *Jensen's inequality.*

Proposition 2.6 (Jensen's inequality, [GG18, Proposition 5.10.3])**.** *Assume a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a random variable $X : \Omega \to \mathbb{R}$. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function, that is,*

$$\varphi(tx + (1-t)y) \le t\varphi(x) + (1-t)\varphi(y)$$

for all $x, y \in \mathbb{R}$ and $t \in [0, 1]$. Then

$$\varphi(\mathbb{E}X) \le \mathbb{E}\left[\varphi(X)\right]. \tag{2.1}$$

If $\psi : \mathbb{R} \to \mathbb{R}$ is concave, that is, $-\psi$ is convex, then

$$\psi(\mathbb{E}X) \ge \mathbb{E}\left[\psi(X)\right]. \tag{2.2}$$

Remark 2.7. The equation (2.2) follows from (2.1) by choosing $\varphi = -\psi$.

Next we consider the following basic form of Fubini's theorem. For the definition of the product of measure spaces, see [GG18, Section 4.3].

Proposition 2.8 ([GG18, Theorem 5.7.3])**.** *Assume two measure spaces* $(\Omega_1, \mathcal{F}_1, \mu_1)$ *and* $(\Omega_2, \mathcal{F}_2, \mu_2)$*. Consider the product space* $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2, \mu_1 \otimes \mu_2)$*. Let* $f : \Omega_1 \times \Omega_2 \to \mathbb{R}$ *be a measurable function. Assume that*

$$\int_{\Omega_1 \times \Omega_2} |f(\omega_1, \omega_2)| \, \mathrm{d}(\mu_1 \otimes \mu_2)(\omega_1, \omega_2) < \infty.$$

Then

$$\begin{aligned}\int_{\Omega_1 \times \Omega_2} f(\omega_1, \omega_2) \, \mathrm{d}(\mu_1 \otimes \mu_2)(\omega_1, \omega_2) &= \int_{\Omega_1} \left[\int_{\Omega_2} f(\omega_1, \omega_2) \, \mathrm{d}\mu_2(\omega_2) \right] \mathrm{d}\mu_1(\omega_1) \\ &= \int_{\Omega_2} \left[\int_{\Omega_1} f(\omega_1, \omega_2) \, \mathrm{d}\mu_1(\omega_1) \right] \mathrm{d}\mu_2(\omega_2).\end{aligned}$$

Fubini's theorem has the following application, which we use throughout this thesis. Let $(\Omega, \mathcal{F}, \mu)$ be a probability space and denote by λ the Lebesgue measure as earlier. If we have a product space $(\Omega \times \mathbb{R}, \mathcal{F} \otimes \mathcal{B}(\mathbb{R}), \mathbb{P} \otimes \lambda)$ and a measurable function $f : \Omega \times [a, b] \to \mathbb{R}$ such that the map $t \mapsto f(\omega, t)$ is continuous for all $\omega \in \Omega$, then

$$\begin{aligned}\mathbb{E} \int_a^b f(\cdot, t) \, \mathrm{d}t &= \int_{\Omega} \int_{[a,b]} f(\omega, t) \, \mathrm{d}\lambda(t) \, \mathrm{d}\mathbb{P}(\omega) \\ &= \int_{[a,b]} \int_{\Omega} f(\omega, t) \, \mathrm{d}\mathbb{P}(\omega) \, \mathrm{d}\lambda(t) \\ &= \int_a^b \mathbb{E} \left[f(\cdot, t) \right] \mathrm{d}t,\end{aligned}$$

provided that $\mathbb{E} \int_a^b |f(\cdot, t)| \, \mathrm{d}t < \infty$.

In the next two theorems we assume a measure space $(\Omega, \mathcal{F}, \mu)$. The first theorem is known as *Hölder's inequality*, and it is one of the most essential inequalities in measure theory.

Proposition 2.9 (Hölder's inequality, [Rud70, Theorem 3.8])**.** *Let* $p, q \in (1, \infty)$ *with* $\frac{1}{p} + \frac{1}{q} = 1$*. Assume real-valued measurable functions* $f \in \mathcal{L}_p(\Omega, \mathcal{F}, \mu)$ *and* $g \in \mathcal{L}_q(\Omega, \mathcal{F}, \mu)$*. Then*

$$\int_{\Omega} |fg| \, \mathrm{d}\mu \leq \left(\int_{\Omega} |f|^p \, \mathrm{d}\mu \right)^{\frac{1}{p}} \left(\int_{\Omega} |g|^q \, \mathrm{d}\mu \right)^{\frac{1}{q}}.$$

Another fundamental inequality is *Minkowski inequality*, which implies the triangle inequality in the $L_p(\Omega, E)$-spaces, which we define in Appendix A.

Proposition 2.10 (Minkowski's inequality, cf. [Rud70, Theorem 3.9]). *Let $(E, \|\cdot\|)$ be a separable Banach space. Assume $p \in [1, \infty)$ and measurable maps $f, g \in L_p(\Omega, E)$. Then*

$$\int_{\mathbb{R}} \|f+g\|_E^p \,\mathrm{d}\mu \leq \left(\int_{\mathbb{R}} \|f\|_E^p \,\mathrm{d}\mu)\right)^{\frac{1}{p}} + \left(\int_{\mathbb{R}} \|g\|_E^p \,\mathrm{d}\mu\right)^{\frac{1}{p}}.$$

Remark 2.11. In [Rud70, Theorem 3.9] Minkowski inequality is proven for real-valued measurable functions. However, it follows in our case by using triangle inequality and noticing

$$\int_{\Omega} \|f+g\|_E^p \,\mathrm{d}\mu \leq \int_{\Omega} |\|f\|_E + \|g\|_E|^p \,\mathrm{d}\mu.$$

2.2.5 Convergence of random variables

There exists several types of convergence of random variables with certain connections. We introduce here those, which we need in this thesis. We follow [GG18, chapter 6].

Let $(E, \|\cdot\|_E)$ be a separable Banach space. We assume a sequence of random variables $(f_n)_{n=1}^{\infty}$, $f_n : \Omega \to E$ and a measurable map $f : \Omega \to E$.

Definition 2.12 (cf. [GG18, definition 6.1.1]). The sequence $(f_n)_{n=1}^{\infty}$ converges *almost surely* to the limit f, if

$$\mathbb{P}(\{\omega \in \Omega \mid \|f_n - f\|_E \to 0 \textit{ as } n \to \infty\}) = 1.$$

We denote this convergence by $f_n \underset{a.s.}{\longrightarrow} f$.

Definition 2.13 (cf. [GG18, definition 6.2.2]). The sequence $(f_n)_{n=1}^{\infty}$ converges to the limit f *in probability*, if for all $\epsilon > 0$ one has

$$\lim_{n\to\infty} \mathbb{P}(\{\omega \in \Omega \mid \|f_n(\omega) - f(\omega)\|_E > \epsilon\}) = 0.$$

We denote this convergence by $f_n \underset{\mathbb{P}}{\longrightarrow} f$.

Almost sure convergence implies convergence in probability.

Proposition 2.14 (cf. [GG18, Proposition 6.2.4 (1)]). *Assume that $(f_n)_{n=1}^{\infty}$ converges to f almost surely. Then f_n converges to f in probability.*

Definition 2.15 (cf. [GG18, definition 6.3.1])**.** Let $p \in (0, \infty)$. Assume that $f_n \in L_p(\Omega, E)$ for all $n \in \mathbb{N}$. The sequence $(f_n)_{n=1}^{\infty}$ converges to the limit $f \in L_p(\Omega, E)$ *with respect to the p-th mean*, if

$$\mathbb{E} \|f_n - f\|_E^p \to 0$$

as n tends to ∞. We denote this convergence by $f_n \xrightarrow[L_p]{} f$.

Under certain conditions the converge in probability implies converge with respect to the p-th mean.

Proposition 2.16 (cf. [GG18, Proposition 6.3.2 (4)])**.** *Let $p \in (0, \infty)$. Assume that $(f_n)_{n=1}^{\infty}$ converges to f in probability. If*

$$\mathbb{E} \sup_{n \in \mathbb{N}} \|f_n\|_E^p < \infty,$$

then $f \in L_p(\Omega, E)$ and $f_n \xrightarrow[L_p]{} f$.

2.3 Stochastic processes

In this section we focus on stochastic processes, which is a necessary component in the theory of stochastic differential equations. We give the basic definitions and the most essential results we need later on in this thesis. In this section we follow [Gei19, chapter 2] and [Mao07, Section 1.3]

Assume a stochastic basis $(\Omega, \mathcal{F}, \mathbb{P}, (F_t)_{t \in I})$. A family of $(\mathcal{F}, \mathcal{B}(\mathbb{R}^d))$-measurable random variables $(X_t)_{t \in I}$ is called a *(stochastic) process*, if $X_t : \Omega \to \mathbb{R}^d$ is a random variable for all $t \in I$.

The measurability of stochastic processes can be classified in the following way, according to [Mao07, p. 10] and [Gei19, definition 2.1.9]:

(1) The process X is *adapted*, if the random variable X_t is $\mathcal{F}_t$-measurable for all $t \in I$.

(2) The process X is *measurable*, if the map $\varphi : \Omega \times I \to \mathbb{R}^d$, $\varphi(\omega, t) := X_t(\omega)$, is $(\mathcal{F}_s \otimes \mathcal{B}(I), \mathcal{B}(\mathbb{R}^d)$-measurable.

(3) The process X is *progressively measurable* with respect to the filtration $(\mathcal{F}_t)_{t \in I}$, if for all $S \in I$ the map $\varphi : \Omega \times [0, S] \to \mathbb{R}^d$, $\varphi(\omega, s) := X_s(\omega)$ is $(\mathcal{F}_S \otimes \mathcal{B}([0, S], \mathcal{B}(\mathbb{R}^d)$-measurable.

Moreover, we say that the process X is (path-wise) continuous, if for all $\omega \in \Omega$ the trajectory $t \mapsto X_t(\omega)$ is continuous.

The following definitions are mentioned in [Mao07, p. 10-11]. If we have two processes $X = (X_t)_{t \in I}$ and $Y = (Y_t)_{t \in I}$ with respect to the same stochastic basis, then we say that X and Y are *indistinguishable* provided that the set

$$\{X_t = Y_t, t \in I\} = \{\omega \in \Omega \mid X_t(\omega) = Y_t(\omega), t \in I\}$$

is measurable and

$$\mathbb{P}(X_t = Y_t, t \in I) = 1.$$

If we have

$$\mathbb{P}(X_t = Y_t) = 1,$$

for all $t \in I$, then X and Y are called *modifications of each other.* It is clear that if two processes are indistinguishable, they are also modifications of each other. The converse implication does not hold in general. However, we have the following proposition.

Proposition 2.17 ([Gei19, Proposition 2.1.7])**.** *Assume two processes $X = (X_t)_{t \in I}$ and $Y = (Y_t)_{t \in I}$ that are modifications of each other. If all the trajectories of X and Y are continuous, then the processes X and Y are indistinguishable.*

2.3.1 Martingales

A special subset of stochastic processes are martingales. Assume a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{G} \subseteq \mathcal{F}$ be a sub-σ-algebra of $\mathcal{F}$, that is, $\mathcal{G}$ is a σ-algebra such that it is also a subset of $\mathcal{F}$. Assume a $\mathcal{G}$-measurable random variable $f : \Omega \to \mathbb{R}^d$ with $\mathbb{E}\,\|f\| < \infty$. The *conditional expectation* of f given $\mathcal{G}$ is a $\mathcal{G}$-measurable random variable $g : \Omega \to \mathbb{R}^d$ satisfying $\mathbb{E}\,\|g\| < \infty$ and

$$\int_B f \,\mathrm{d}\mathbb{P} = \int_B g \,\mathrm{d}\mathbb{P}$$

for all $B \in \mathcal{G}$. We denote

$$\mathbb{E}\,[f \mid \mathcal{G}] := g.$$

The conditional expectation is almost surely unique, meaning that if there exists another g' having the same properties as g, then we have that $\mathbb{P}(g = g') = 1$.

Next assume a stochastic process $X = (X_t)_{t \in I}$. The process X is a martingale, provided that the following two conditions are satisfied.

(1) For all $t \in I$ it holds that X_t is $\mathcal{F}_t$-measurable and $\mathbb{E}\,\|X_t\| < \infty$.

(2) For all $s, t \in I$ with $s < t$ it holds that

$$\mathbb{E}[X_t \mid \mathcal{F}_s] = X_s$$

almost surely, that is,

$$\mathbb{P}(\{\omega \in \Omega \mid \mathbb{E}[X_t \mid \mathcal{F}_s](\omega) = X_s(\omega)\}) = 1.$$

We use the following notation for the set of martingales:

(1) $\mathcal{M}$: the set of martingales.

(2) $\mathcal{M}^c$: martingales with continuous trajectories, that is, $t \mapsto X_t(\omega)$ is continuous for all $\omega \in \Omega$.

(3) $\mathcal{M}^{c,0}$: continuous martingales with $M_0 \equiv 0$.

(4) $\mathcal{M}_2^{c,0}$: the set of square integrable martingales, that is, if $M \in \mathcal{M}^{c,0}$ and $\mathbb{E}\|M_t\|^2 < \infty$ for all $t \in I$, then $M \in \mathcal{M}_2^{c,0}$.

If it is not clear from the context, we may write $\mathcal{M}(\mathbb{R}^d)$ to emphasize the dimension.

A martingale $X = (X_t)_{t\in I} \in \mathcal{M}$ has the property $\mathbb{E}X_t = \mathbb{E}X_0$ for all $t \in I$. In particular, if $X \in \mathcal{M}^{c,0}$, then $\mathbb{E}X_t = 0$ for all $t \in I$.

2.3.2 Brownian motion

Here we follow [Gei19, definition 2.4.5]. Assume a stochastic basis $(\Omega, \mathcal{F}, \mathbb{P}, (F_t)_{t\in I})$ that satisfies the usual conditions. Let $B = (B_t)_{t\in I}$ be an adapted process, that is, B_t is $\mathcal{F}_t$-measurable for all $t \in I$. The process B is called *(standard) Brownian motion* with respect to $(\mathcal{F}_i)_{i\in I}$ provided that the following condition are satisfied.

(1) $B_0 \equiv 0$.

(2) $B_t - B_s$ is independent from $\mathcal{F}_s$ for all $s, t \in I$, $s < t$, meaning that

$$\mathbb{P}(C \cap \{B_t - B_s \in A\}) = \mathbb{P}(C)\mathbb{P}(B_t - B_s \in A)$$

for all $C \in \mathcal{F}_s$ and $A \in \mathcal{B}(\mathbb{R})$.

(3) $B_t - B_s \sim \mathcal{N}(0, t - s)$ for all $s, t \in I$, $s < t$.

(4) The trajectories $t \mapsto B_t(\omega)$ are continuous for all $\omega \in \Omega$.

If $B^i = ({B_t}^i)_{t\in I}$ is a Brownian motion for all $i = 1, ..., d$ and the Brownian motions $B_t^1, ..., B_t^d$ are independent from each other, $B = (B^1, B^2, ..., B^d)$ is called *d-dimensional Brownian motion.*

2.4 Stochastic calculus

Stochastic calculus includes, amongst other parts, the theory of stochastic integration and stochastic differential equations. In this section we have two objectives. The first one is to give a proper definition for a stochastic integral with respect to the Brownian motion. The second objective is to define ordinary stochastic differential equations and give known existence and uniqueness results for them, so we can later generalize these results to a wider class of stochastic differential equations.

2.4.1 Stochastic integration

In this section we introduce stochastic integration with respect to a Brownian motion. We start in the one-dimensional case, and then generalize the definition to multiple dimensions. We follow [Gei19, chapter 3] and [Mao07, Section 1.5].

We start by defining the stochastic integral for simple processes. The definition of a simple process is given in [Gei19, definition 3.1.1]. Assume a stochastic basis $(\Omega, \mathcal{F}, \mathbb{P}, (F_t)_{t \in I})$ that satisfies the usual conditions. Let $B = (B_t)_{t \in I}$ be a one dimensional $(F_t)_{t \in I}$ Brownian motion.

A real-valued stochastic process $L = (L_t)_{t \in I}$ is called *simple*, if there exists a finite sequence $(t_k)_{k=1}^n$ of real numbers satisfying

$$0 = t_0 < t_1 < t_2 < ... < t_n = T$$

and $(\mathcal{F}_{t_i}, \mathcal{B}(\mathbb{R}))$-measurable random variables $v_i : \Omega \to \mathbb{R}$, $i \in \mathbb{N}$, with

$$\sup_{(i,\omega) \in \mathbb{N} \times \Omega} |v_i(\omega)| < \infty,$$

such that

$$L_t(\omega) = \sum_{i=1}^{\infty} \mathbb{1}_{(t_{i-1}, t_i]}(t) v_{i-1}(\omega).$$

We denote by $\mathcal{L}_0(\mathbb{R})$ the space of simple processes.

Next we define the stochastic integral for simple processes.

Definition 2.18 ([Gei19, definition 3.1.2])**.** Let $L \in \mathcal{L}_0(\mathbb{R})$. The stochastic integral for $\mathcal{L}_0(\mathbb{R})$ integrand L with respect to the Brownian motion B is defined by

$$I_t(L)(\omega) := \sum_{k=1}^{\infty} v_{k-1}(\omega)(B_{t_k \wedge t}(\omega) - B_{t_{k-1} \wedge t}(\omega)).$$

By [Gei19, proposition 3.1.6], it holds that $I_t(L) \in \mathcal{M}_2^{c,0}$.

Stochastic integral can be expanded to a much larger space. We denote by $\mathcal{L}_2(\mathbb{R})$ all the processes $L = (L_t)_{\in I}$ that are progressively measurable and satisfy the property

$$\mathbb{E} \int_0^t |L_u|^2 \, \mathrm{d}u < \infty$$

for all $t \in I$. We see that $\mathcal{L}_0(\mathbb{R}) \subset \mathcal{L}_2(\mathbb{R})$.

The next theorem provides a way to generalize the stochastic integral to the set $\mathcal{L}_2(\mathbb{R})$.

Theorem 2.19 ([Gei19, Proposition 3.1.12])**.** *The map $I : \mathcal{L}_0(\mathbb{R}) \to \mathcal{M}_2^{c,0}$ can be generalized to a map $J : \mathcal{L}_2(\mathbb{R}) \to \mathcal{M}_2^{c,0}$ such that the following properties are satisfied:*

(1) For $\alpha, \beta \in \mathbb{R}$ and $K, L \in \mathcal{L}_2(\mathbb{R})$ one has that

$$J_t(\alpha K + \beta L) = \alpha J_t(K) + \beta J_t(L)$$

for $t \in I$ almost surely.

(2) If $L \in \mathcal{L}_0(\mathbb{R})$, then $I_t(L) = J_t(L)$ for $t \in I$ almost surely.

(3) If $L \in \mathcal{L}_2(\mathbb{R})$, then

$$\left(\mathbb{E} \, |J_t(L)|^2\right)^{\frac{1}{2}} = \left(\mathbb{E} \int_0^t L_u^2 \, \mathrm{d}u\right)^{\frac{1}{2}}$$

for $t \in I$.

(4) If $L \in \mathcal{L}_2(\mathbb{R})$ and $(A_n)_{n=1}^{\infty}$ is a sequence of processes in $\mathcal{L}_2(\mathbb{R})$ such that $d(A_n, L) \to 0$ as $n \to \infty$, then

$$\mathbb{E} \sup_{t \in I} |J_t(L) - J_t(A_n)|^2 \to 0$$

as $n \to \infty$.

(5) If J' is another map satisfying the properties above, then

$$\mathbb{P}(J_t(L) = J_t'(L) \textit{ for all } t \in I) = 1$$

for all $L \in \mathcal{L}_2(\mathbb{R})$.

Definition 2.20. The process $X_t := J_t(L)$, $L \in \mathcal{L}_2(\mathbb{R})$, obtained in 2.19 is called the *stochastic integral* of L with respect to B until time t, and we denote

$$X_t = \int_0^t L_s \, \mathrm{d}B_s.$$

Since $\int_0^t L_s \,\mathrm{d}B_s \in \mathcal{M}_2^{c,0}$, one has that

$$\mathbb{E}\int_0^t L_s \,\mathrm{d}B_s = 0.$$

To compute the second moment, one can use a theorem known as *Itô's isometry.*

Proposition 2.21 (Itô's isometry, [Gei19, 3.1.25 (iii)])**.** *Let* $L \in \mathcal{L}_2(\mathbb{R})$. *Then*

$$\mathbb{E}\left[\left(\int_0^t L_s \,\mathrm{d}B_s\right)^2\right] = \mathbb{E}\left[\int_0^t L_s^2 \,\mathrm{d}s\right].$$

The stochastic integral can be generalized to multiple dimensions in a simple way. If $X = (X_t)_{t\in I} = \left(\left[X_t^{ij}\right]\right)_{t\in I}$ is an $\mathbb{R}^{d\times m}$-valued process such that $X^{ij} = (X_t^{ij})_{t\in I} \in \mathcal{L}_2(\mathbb{R})$ for all $i = 1,...,d, j = 1,...,m$, then we write that $X \in \mathcal{L}_2(\mathbb{R}^{d\times m})$.

Definition 2.22 ([Mao07, Section 1.5, Definition 5.20])**.** Let $d, m \in \mathbb{N}$ and let $B = (B^1,...B^m)$ be an m-dimensional Brownian motion. Assume an $\mathbb{R}^{d\times m}$-valued process $X \in \mathcal{L}_2(\mathbb{R}^{d\times m})$. We define

$$\int_0^t X_s \,\mathrm{d}B_s = \int_0^t \begin{bmatrix} X_s^{11} & \dots & X_s^{1d} \\ \vdots & \ddots & \vdots \\ X_s^{m1} & \dots & X_s^{md} \end{bmatrix} \begin{bmatrix} \mathrm{d}B_s^1 \\ \vdots \\ \mathrm{d}B_s^m \end{bmatrix} =: \begin{bmatrix} A_1 \\ \vdots \\ A_d \end{bmatrix}, \tag{2.3}$$

where

$$A_i = \sum_{j=1}^m \int_0^t X_s^{ij} \,\mathrm{d}B_s^j$$

for all $i = 1,...,d$.

It should be noted that the extension to multiple dimensions preserves the martingale property, that is,

$$\int_0^t X_s \,\mathrm{d}B_s \in \mathcal{M}_2^{c,0}(\mathbb{R}^d)$$

for all $X \in \mathcal{L}_2(\mathbb{R}^d)$, which follows from that each component of (2.3) is a finite sum of one-dimensional stochastic integrals, which are in $\mathcal{M}_2^{c,0}(\mathbb{R})$ as we have stated earlier.

The following theorem is known as the *Burkholder-Davis-Gundy inequality*, which can be used to estimate the norms of stochastic integrals.

Proposition 2.23 (Burkholder-Davis-Gundy, [Mao07, Section 1.7, theorem 7.3])**.** *Let $L \in \mathcal{L}_2(\mathbb{R}^{d\times m}$ and $p \in (0,\infty)$. Then there exist constants $c_p, C_p > 0$ depending only on p such that*

$$c_p\mathbb{E}\left[\sqrt{\int_0^T \|L_s\|^2 \,\mathrm{d}s}\right]^p \leq \mathbb{E}\left(\sup_{t\in[0,T]}\left\|\int_0^t L_s \,\mathrm{d}B_s\right\|^p\right) \leq C_p\mathbb{E}\left[\sqrt{\int_0^T \|L_s\|^2 \,\mathrm{d}s}\right]^p.$$

2.4.2 Itô's formula

A powerful tool in classical stochastic calculus is *Itô's formula.* It can be used, for example, to find explicit formulas for stochastic integrals and to solve ordinary stochastic differential equations. We follow [Gei19, chapter 3-4]. We only consider the one-dimensional case. A multidimensional version can be found in [Mao07, Section 1.3, theorem 3.4]. The following definition of Itô process is given in [Gei19, definition 3.2.6].

We recall that a continuous and adapted process $X = (X_t)_{t\in I}$, $X_t : \Omega \to \mathbb{R}$, is called *Itô process*, provided that there exists $x_0 \in \mathbb{R}$, a process $L = (L_t)_{t\in I} \in \mathcal{L}_2$ and a progressively measurable process $a = (a_i)_{i\in I}$ with

$$\int_0^t |a_u(\omega)| \,\mathrm{d}u < \infty$$

for all $(t,\omega) \in I \times \Omega$, such that

$$X_t(\omega) = x_0 + \left(\int_0^t L_u \,\mathrm{d}B_u\right)(\omega) + \int_0^t a_u(\omega) \,\mathrm{d}u$$

for $t \in I$ almost surely.

We say that X is an Itô process with representation (x_0, L, a).

Theorem 2.24 (Itô's formula, [Gei19, Proposition 3.2.9])**.** *Let $X = (X_t)_{t\in I}$ be an Itô process with representation (x_0, L, a) and let $f \in C^{1,2}(I \times \mathbb{R})$. Then*

$$\begin{aligned} f(t, X_t) = f(0, X_0) &+ \int_0^t \frac{\partial f}{\partial u}(u, X_u) \,\mathrm{d}u + \int_0^t \frac{\partial f}{\partial x}(u, X_u) \,\mathrm{d}X_u \\ &+ \frac{1}{2}\int_0^t \frac{\partial^2 f}{\partial x^2}(u, X_u) L_u^2 \,\mathrm{d}u, \end{aligned}$$

where

$$\int_0^t \frac{\partial f}{\partial x}(u, X_u) \,\mathrm{d}X_u = \int_0^t \frac{\partial f}{\partial x}(u, X_u) L_u \,\mathrm{d}B_u + \int_0^t \frac{\partial f}{\partial x}(u, X_u) a_u \,\mathrm{d}u,$$

for $t \in I$ almost surely.

2.4.3 Stochastic differential equations

Next we introduce ordinary stochastic differential equations (SDE), where the coefficient functions depend only on the time variable and the unknown process on a certain time. Despite the name, stochastic differential equations are more related to integral equations than classical differential equations. We follow [Gei19, chapter 4] and [Mao07, chapter 2].

Assume a stochastic basis $(\Omega, \mathcal{F}, \mathbb{P}, (F_t)_{t\in I})$ that satisfies the usual conditions. We denote by $B = (B_t)_{t\in I}$ a d-dimensional $(\mathcal{F}_t)_{t\in I}$ Brownian motion.

Definition 2.25 ([Gei19, definition 4.1.1], [Mao07, sector 2.2, definition 2.1])**.** Assume that the coefficients

$$b : I \times \mathbb{R}^d \to \mathbb{R}^d$$

and

$$\sigma : I \times \mathbb{R}^d \to \mathbb{R}^{d\times d},$$

are Borel measurable. Let $x_0 \in \mathbb{R}^d$ and assume an open set $D \subseteq \mathbb{R}^d$. An adapted and path-wise continuous process $X = (X_t)_{t\in I}$ solves the *stochastic differential equation*

$$\begin{cases} \mathrm{d}X_t = \sigma(t, X_t)\,\mathrm{d}B_t + b(t, X_t)\,\mathrm{d}t \\ X_0 = x \end{cases} \tag{2.4}$$

if the following conditions are satisfied:

(1) $X_t(\omega) \in D$ for all $t \in I$ and $\omega \in \Omega$.

(2) $X_0 \equiv x_0$.

(3)
$$X_t = x_0 + \int_0^t \sigma(s, X_s)\,\mathrm{d}B_s + \left(\int_0^t b(s, X_s)_1\,\mathrm{d}s, ..., \int_0^t b(s, X_s)_d\,\mathrm{d}s\right),$$
where $X_s = (X_s^1, ..., X_s^d)$, for all $t \in I$ almost surely.

We call the term $\sigma(t, X_t)\,\mathrm{d}B_t$ the *diffusion term* and the term $b(t, X_t)\,\mathrm{d}t$ the *drift term*. Respectively, σ and b are called diffusion and drift coefficients.

Remark 2.26. In Definition 2.25 Borel-measurability means that σ is $(\mathcal{B}(I)\otimes\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^{d\times d}))$-measurable and b is $(\mathcal{B}(I)\otimes\mathcal{B}(\mathbb{R}^d), \mathcal{B}(\mathbb{R}^d))$-measurable.

2.4.4 Existence and uniqueness of a solution

Next we formulate two theorems about the existence and uniqueness of a solution. However, first we need to define what we mean by uniqueness.

Assume any two processes $X = (X_t)_{t\in I}$ and $Y = (Y_t)_{t\in I}$ that solve the SDE (2.4). If X and Y are indistinguishable, that is,

$$\mathbb{P}(X_t = Y_t, t \in I) = 1,$$

then it is said that the SDE (2.4) has a *unique* strong solution.

Our first existence and uniqueness theorem is usually referred as *existence under Lipschitz condition*, although alongside global Lipschitz condition we also assume that the coefficient functions satisfy linear growth condition, which is necessary to make sure the coefficients do not grow too fast.

Theorem 2.27 ([Mao07, Section 2.3, theorem 3.1, lemma 3.2]). *Suppose that the coefficient functions σ and b are continuous and there exists a constant $C > 0$ such that*

(C1)

$$\|b(t,x)\| + \|\sigma(t,x)\| \leq C\,(1 + \|x\|)$$

for all $t \in I$ and $x \in \mathbb{R}^d$, and

(C2)

$$\|b(t,x) - b(t,y)\| + \|\sigma(t,x) - \sigma(t,y)\| \leq C\,\|x - y\|$$

for all $t \in I$ and $x, y \in \mathbb{R}^d$.

Under these conditions the SDE (2.4) *admits a unique strong solution.*

Our second theorem gives the uniqueness of a solution under weaker conditions than our previous theorem. However, it should be noted that this theorem does not imply the existence of a solution, just the uniqueness, and the theorem is given in the one-dimensional case.

Theorem 2.28 (Yamada-Tanaka, [Gei19, Proposition 4.2.3]). *Let $h : [0,\infty) \to [0,\infty)$ and $K : [0,\infty) \to \mathbb{R}$ be strictly increasing functions such that $K(0) = h(0) = 0$, K is concave, and for all $\epsilon > 0$ it holds that*

$$\int_0^\epsilon \frac{1}{K(u)}\,\mathrm{d}u = \int_0^\epsilon \frac{1}{h(u)^2}\,\mathrm{d}u = \infty.$$

If the coefficient functions σ and b are continuous and

$$|\sigma(t,x) - \sigma(t,y)| \leq h(|x-y|),$$
$$|b(t,x) - b(t,y)| \leq K(|x-y|)$$

for all $x, y \in \mathbb{R}$, then any two solutions to the SDE (2.4) *are indistinguishable.*

3 Wasserstein space

In this thesis our primary objective is to generalize the theory of ordinary stochastic differential equations to a wider class of equations, where the coefficients may depend upon the law of the unknown process. However, we need to address the following problems:

(1) The coefficients functions σ and b are required to be Borel measurable. If $X : \Omega \to \mathbb{R}^d$ is a random variable, then its law $\mathbb{P}_X$ is a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. It follows that we need to define a Borel σ-algebra on the space of probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

(2) If we want to formulate the existence and uniqueness Theorem 2.27 for this wider class of stochastic differential equations, we may need to define Lipschitz-continuity with respect to the distribution variable, that is, for all probability measures μ, ν on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ one has

$$\|b(t, x, \mu) - b(t, x, \nu)\| \leq d(\mu, \nu),$$

where d is a distance between two probability measures.

For this purpose we introduce the *Wasserstein distance*, which is a metric for the space of probability measures with finite p-th moments. In this section we define the Wasserstein space and prove some important properties.

The definition of marginal distributions follow [Vil06, chapter 1]. The space $\mathcal{P}_p(\mathbb{R}^d)$ and the Wasserstein distance W_p are defined in [Vil06, definition 6.1 and 6.4].

Let X be a non-empty set. A map $d : X \times X \to [0, \infty)$ is a *metric* or *distance* if for all $x, y, z \in X$ one has

(M1) $d(x, y) = 0$ if and only if $x = y$.

(M2) $d(x, y) = d(y, x)$.

(M3) $d(x, y) \leq d(x, z) + d(z, y)$.

The pair (X, d) forms a *metric space*. One important example of a metric space is $\mathbb{R}^d$ with Euclidean metric $d_E(x, y) := \|x - y\|$, where $\|\cdot\|$ is the ordinary Euclidean norm. In particular this space is complete and separable.

Let $\mathcal{P}(\mathbb{R}^d)$ be the space of all the probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. For $p \geq 1$, let $\mathcal{P}_p(\mathbb{R}^d)$ be a subspace of $\mathcal{P}(\mathbb{R}^d)$ such that

$$\mathcal{P}_p(\mathbb{R}^d) := \left\{ \mathbb{P} \in \mathcal{P}(\mathbb{R}^d) \mid \int_{\mathbb{R}} \|x - x_0\|^p \, d\mathbb{P} < \infty \right\},$$

where $x_0 \in \mathbb{R}^d$ is fixed.

Denote by $\Pi(\mu, \nu)$ the set of probability measures on $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d))$ where the first and second marginals are μ and ν respectively. This means $\xi \in \Pi(\mu, \nu)$, if

(1) ξ is a measure on $(\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d))$,

(2) $\xi(\mathbb{R}^d \times \mathbb{R}^d) = 1$,

(3) for all $A \in \mathcal{B}(\mathbb{R}^d)$ one has

$$\mu(A) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{1}_A(x) \, \mathrm{d}\xi(x, y) = \xi(A \times \mathbb{R}^d)$$

and

$$\nu(A) = \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathbb{1}_A(y) \, \mathrm{d}\xi(x, y) = \xi(\mathbb{R}^d \times A).$$

Example 3.1. Let $X, Y : \Omega \to \mathbb{R}^d$ be random variables. The law of the random vector (X, Y) is defined by

$$\mathbb{P}_{(X,Y)}(B) := \mathbb{P}((X, Y) \in B)$$

for all $B \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d)$. For $A \in \mathcal{B}(\mathbb{R}^d)$ we have

$$\mathbb{P}_{(X,Y)}(A \times \mathbb{R}^d) = \mathbb{P}((X, Y) \in A \times \mathbb{R}^d) = \mathbb{P}(X \in A) = \mathbb{P}_X(A)$$

and in a similar way $\mathbb{P}_{(X,Y)}(\mathbb{R}^d \times A) = \mathbb{P}_Y(A)$. It follows that

$$\mathbb{P}_{(X,Y)} \in \Pi(\mathbb{P}_X, \mathbb{P}_Y).$$

Definition 3.2 ([Vil06, definition 6.1 and 6.4])**.** For all $p \geq 1$, define $W_p : \mathcal{P}_p(\mathbb{R}^d) \times \mathcal{P}_p(\mathbb{R}^d) \to [0, \infty)$,

$$W_p(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p \, \mathrm{d}\pi(x, y) \right)^{\frac{1}{p}}.$$

The map W_p is called the *p-Wasserstein distance*. The space $(\mathcal{P}_p(\mathbb{R}^d), W_p)$ is called the *Wasserstein space*.

Theorem 3.3. *Let $p \geq 1$. Then the Wasserstein space $\big(\mathcal{P}_p(\mathbb{R}^d), W_p\big)$ is a complete and separable metric space.*

Proof. First we need to show that W_p satisfies properties (**M1**), (**M2**) and (**M3**). The triangle inequality property (**M3**) is proven in [Vil06, chapter 6, p. 77]. The remaining parts we prove here.

To prove the symmetry property (**M2**) we let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$. For all $A \in \mathcal{B}(\mathbb{R}^d)$, $\pi \in \Pi(\mu,\nu)$ and $\xi \in \Pi(\nu,\mu)$ one has

$$\pi(A \times \mathbb{R}^d) = \mu(A) = \xi(\mathbb{R}^d \times A) \quad \text{and} \quad \pi(\mathbb{R}^d \times A) = \nu(A) = \xi(A \times \mathbb{R}^d).$$

Now we may define a map $\rho : \mathcal{P}_p(\mathbb{R}^d \times \mathbb{R}^d) \to \mathcal{P}_p(\mathbb{R}^d \times \mathbb{R}^d)$ such that for all $\pi \in \mathcal{P}_p(\mathbb{R}^d \times \mathbb{R}^d)$ and all $B \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d)$ one has

$$(\rho(\pi))(B) = \pi(\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d \mid (y,x) \in A\}) = \xi(B).$$

In particular $\rho(\Pi(\mu,\nu)) = \Pi(\nu,\mu)$. We see that $\rho^{-1} = \rho$ because $\rho(\rho(\pi)) = \pi$ for all $\pi \in \mathcal{P}_p(\mathbb{R}^d \times \mathbb{R}^d)$. Hence $\rho^{-1}(\Pi(\nu,\mu)) = \rho(\Pi(\nu,\mu)) = \Pi(\mu,\nu)$. Now

$$\begin{aligned}
W_p(\mu,\nu)^p &= \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x-y\|^p \,\mathrm{d}\pi(x,y) \\
&= \inf_{\pi \in \rho(\Pi(\nu,\mu))} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x-y\|^p \,\mathrm{d}\pi(x,y) \\
&= \inf_{\rho^{-1}(\pi) \in \Pi(\nu,\mu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x-y\|^p \,\mathrm{d}\pi(x,y) \\
&= \inf_{\xi \in \Pi(\nu,\mu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x-y\|^p \,\mathrm{d}\rho(\xi)(x,y) \\
&= \inf_{\xi \in \Pi(\nu,\mu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|y-x\|^p \,\mathrm{d}\xi(y,x) \\
&= W_p(\nu,\mu)^p.
\end{aligned}$$

We prove the final property (**M1**) in two steps. First we prove that $\mu = \nu$ implies that $W_p(\mu,\nu) = 0$. We define a measure

$$\pi_0(B) := \int_{\mathbb{R}^d} \mathbb{1}_{\{x \in \mathbb{R}^d \mid (x,x) \in B\}}(y) \,\mathrm{d}\mu(y)$$

for $B \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d)$. Clearly,

$$\begin{aligned}
\pi_0(A \times \mathbb{R}^d) &= \int_{\mathbb{R}^d} \mathbb{1}_{\{x \in \mathbb{R}^d \mid (x,x) \in A \times \mathbb{R}^d\}}(y) \,\mathrm{d}\mu(y) \\
&= \int_{\mathbb{R}^d} \mathbb{1}_A(y) \,\mathrm{d}\mu(y) = \mu(A).
\end{aligned}$$

The same arguments can be used to show that $\pi_0(\mathbb{R}^d \times A) = \mu(A)$. Hence $\pi_0 \in \Pi(\mu,\mu)$.

We see that for all sets $B \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}^d)$ with $B \cap \{(x,x) \mid x \in \mathbb{R}^d\} = \emptyset$ one has $\pi_0(B) = 0$. Therefore

$$\begin{aligned} W_p(\mu,\mu)^p &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x-y\|^p \,\mathrm{d}\pi_0(x,y) \\ &= \int_{\{(x,x) \mid x \in \mathbb{R}^d\}} \|x-y\|^p \,\mathrm{d}\pi_0(x,y) \\ &= \int_{\{(x,x) \mid x \in \mathbb{R}^d\}} \|x-x\|^p \,\mathrm{d}\pi_0(x,x) = 0. \end{aligned}$$

To prove the converse implication, we let $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ and assume that $W_p(\mu,\nu) = 0$. By [Vil06, Theorem 4.1] this implies that there exists $\pi_0 \in \Pi(\mu,\nu)$ such that

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} \|x-y\|^p \,\mathrm{d}\pi_0(x,y) = 0.$$

Since π_0 is a probability measure, it follows that

$$\pi_0(\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d \mid x = y\}) = 1.$$

In particular

$$\pi_0(\{(x,y) \in \mathbb{R}^d \times \mathbb{R}^d \mid x \neq y\}) = 0.$$

Then for all $A \in \mathcal{B}(\mathbb{R}^d)$ it holds that

$$\begin{aligned} \mu(A) &= \pi_0(A \times \mathbb{R}^d) \\ &= \pi_0(\{(x,y) \in A \times \mathbb{R}^d \mid x = y\} \cup \{(x,y) \in A \times \mathbb{R}^d \mid x \neq y\}) \\ &= \pi_0(\{(x,y) \in A \times \mathbb{R}^d \mid x = y\}) \\ &= \pi_0(A \times A). \end{aligned}$$

With similar arguments we obtain

$$\nu(A) = \pi_0(A \times A).$$

Hence $\mu = \nu$.

In [Vil06, Theorem 6.16] it is proven that if X is a complete separable metric space, then the space $(\mathcal{P}_p(X), W_p)$ is also a complete separable metric space. We use the fact that $\mathbb{R}^d$ with Euclidean metric is complete and separable.

□

We recall some more results concerning the Wasserstein distance. The following lemma and its proof follow [BMM19, Section 2.2].

Lemma 3.4. *Let $X, Y : \Omega \to \mathbb{R}^d$ be random variables. Then*

$$W_p(\mathbb{P}_X, \mathbb{P}_Y)^p \leq \mathbb{E}\left[\|X - Y\|^p\right].$$

Proof. We have shown in example 3.1 that $\mathbb{P}_{(X,Y)} \in \Pi(\mathbb{P}_X, \mathbb{P}_Y)$. Then

$$\begin{aligned} W_p(\mathbb{P}_X, \mathbb{P}_Y)^p &= \left(\inf_{\pi \in \Pi(\mathbb{P}_X, \mathbb{P}_Y)} \left[\int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p \, \mathrm{d}\pi(x, y) \right]^{\frac{1}{p}} \right)^p \\ &= \inf_{\pi \in \Pi(\mathbb{P}_X, \mathbb{P}_Y)} \left[\int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p \, \mathrm{d}\pi(x, y) \right] \\ &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p \, \mathrm{d}\mathbb{P}_{(X,Y)}(x, y). \end{aligned}$$

By letting $\varphi(u, v) := \|u - v\|^p$ we may use change of variable formula [GG18, Proposition 5.6.1] to conclude that

$$\begin{aligned} \mathbb{E}\|X - Y\|^p = \mathbb{E}\varphi(X, Y) &= \int_{\Omega} \varphi(X(\omega), Y(\omega)) \, \mathrm{d}\mathbb{P}(\omega) \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \varphi(x, y) \, \mathrm{d}\mathbb{P}_{(X,Y)}(x, y) \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p \, \mathrm{d}\mathbb{P}_{(X,Y)}(x, y). \end{aligned}$$

Hence

$$W_p(\mathbb{P}_X, \mathbb{P}_Y)^p \leq \mathbb{E}\|X - Y\|^p.$$

□

Due to its complex nature, computing an explicit value for the Wasserstein distance might not be possible. However, in the case $p = 1$, we may apply a theorem known as Kantovich-Rubinstein duality.

Proposition 3.5 (Kantovich-Rubinstein, [CD18a, corollary 5.4])**.** *For $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ one has*

$$W_1(\mu, \nu) = \sup \left\{ \left| \int_{\mathbb{R}^d} h \, \mathrm{d}(\mu - \nu) \right| \mid h \in Lips_1(\mathbb{R}^d) \right\},$$

where $Lips_1(\mathbb{R}^d)$ consists of all the functions $h : \mathbb{R}^d \to \mathbb{R}$ satisfying

$$|h(x) - h(y)| \leq \|x - y\|$$

for all $x, y \in \mathbb{R}^d$.

The following example demonstrates how the Wasserstein distance can be computed in a simple case using Theorem 3.5.

Example 3.6. We define the *Dirac measure* on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ by

$$\delta_c(A) = \begin{cases} 1, & c \in A \\ 0, & c \notin A \end{cases}$$

for some fixed constant $c \in \mathbb{R}$. It is clearly a probability measure. In particular, if we integrate an integrable function f with respect to δ_c, we obtain

$$\int_{\mathbb{R}^d} f \,\mathrm{d}\delta_c = f(c).$$

This implies, for any $p \geq 1$,

$$\int_{\mathbb{R}} \|u\|^p \,\mathrm{d}\delta_c(u) = |c|^p < \infty.$$

Hence $\delta_c \in \mathcal{P}_p(\mathbb{R})$ for all $p \geq 1$.

We let $a, b \in \mathbb{R}^d$. Let $f : \mathbb{R} \to \mathbb{R}$ be a 1-Lipschitz function. Then

$$\begin{aligned} \left| \int_{\mathbb{R}} f \,\mathrm{d}(\delta_a - \delta_b) \right| &= \left| \int_{\mathbb{R}} f \,\mathrm{d}\delta_a - \int_{\mathbb{R}} f \,\mathrm{d}\delta_b \right| \\ &= |f(a) - f(b)| \leq \|a - b\| . \end{aligned}$$

Next we define an orthogonal projection

$$P : \mathbb{R}^d \to \langle b - a \rangle = \left\{ x \in \mathbb{R}^d \mid x = \lambda(b - a) \text{ for some } \lambda \in \mathbb{R} \right\}.$$

It holds that $\|P(x)\| \leq \|x\|$. Now we may let $f(x) = \|P(x - a)\|$ since

$$\begin{aligned} |f(a) - f(b)| &= |\|P(a - a)\| - \|P(b - a)\|| \\ &= |\|P(0)\| - \|P(b - a)\|| \\ &= \|b - a\| . \end{aligned}$$

Furthermore, for all $x, y \in \mathbb{R}^d$ it holds that

$$\begin{aligned} |f(x) - f(y)| &= |\|P(x - a)\| - \|P(y - a)\|| \\ &= |\|P(x - a) - P(y - a)\|| \\ &= \|P(x - y)\| \\ &\leq \|x - y\| , \end{aligned}$$

implying that $f \in \mathrm{Lips}_1 \mathbb{R}^d$.

Now we may apply Theorem 3.5 to conclude that

$$W_1(\delta_a, \delta_b) = \sup_{f \in \mathrm{Lips}_1(\mathbb{R})} \int_{\mathbb{R}} f \,\mathrm{d}(\delta_a - \delta_b) = \|a - b\| .$$

The Wasserstein distances with different p have the following relation. This property is mentioned in [CD18a, p. 353], but it is not proven there.

Lemma 3.7. *Let* $\mu, \nu \in \mathcal{P}_q(\mathbb{R}^d)$*. Then*

$$W_p(\mu, \nu) \leq W_q(\mu, \nu)$$

for all $1 \leq p < q < \infty$.

Proof. Fix $\mu, \nu \in \mathcal{P}_q(\mathbb{R}^d)$ and choose any $\pi \in \Pi(\mu, \nu)$. Let $r = \frac{q}{p}$ and $s = \frac{q}{q-p}$. Now $\frac{1}{r} + \frac{1}{s} = 1$, so we may apply Hölder inequality 2.9 to obtain

$$\begin{aligned}
\int_{\mathbb{R}^d} \|x-y\|^p \, \mathrm{d}\pi(x,y) &\leq \int_{\mathbb{R}^d} |\|x-y\|^p \cdot 1| \, \mathrm{d}\pi(x,y) \\
&\leq \left(\int_{\mathbb{R}^d} 1^s \, \mathrm{d}\pi(x,y) \right)^{\frac{1}{s}} \left(\int_{\mathbb{R}^d} \|x-y\|^{p\frac{q}{p}} \, \mathrm{d}\pi(x,y) \right)^{\frac{1}{r}} \\
&= \left(\int_{\mathbb{R}^d} \|x-y\|^q \, \mathrm{d}\pi(x,y) \right)^{\frac{p}{q}}.
\end{aligned}$$

Hence

$$\left(\int_{\mathbb{R}^d} \|x-y\|^p \, \mathrm{d}\pi(x,y) \right)^{\frac{1}{p}} \leq \left(\int_{\mathbb{R}^d} \|x-y\|^q \, \mathrm{d}\pi(x,y) \right)^{\frac{1}{q}}.$$

Now we have that

$$\inf_{\tilde{\pi} \in \Pi(\mu,\nu)} \left(\int_{\mathbb{R}^d} \|x-y\|^p \, \mathrm{d}\tilde{\pi}(x,y) \right)^{\frac{1}{p}} \leq \left(\int_{\mathbb{R}^d} \|x-y\|^q \, \mathrm{d}\pi(x,y) \right)^{\frac{1}{q}}.$$

This inequality holds for arbitrary $\pi \in \Pi(\mu, \nu)$, therefore

$$\inf_{\tilde{\pi} \in \Pi(\mu,\nu)} \left(\int_{\mathbb{R}^d} \|x-y\|^p \, \mathrm{d}\tilde{\pi}(x,y) \right)^{\frac{1}{p}} \leq \inf_{\pi' \in \Pi(\mu,\nu)} \left(\int_{\mathbb{R}^d} \|x-y\|^q \, \mathrm{d}\pi'(x,y) \right)^{\frac{1}{q}}.$$

□

4 McKean-Vlasov stochastic differential equations

In this thesis we consider a broader class of stochastic differential equations than what we have mentioned in Section 2.4.3. We add a third parameter to the coefficients, a so called *distribution parameter*, which allows us to make the coefficients depend on the law of the random process, and therefore the expected value. This class of stochastic differential equations is called *McKean-Vlasov stochastic differential equations.* We use the abbreviation *MVSDE.* It should be noted that the ordinary SDEs are a special subset of MVSDEs.

Our goal is to generalize some known results of ordinary SDEs to the context of MVSDEs. First we consider theorems for the existence and uniqueness of a solution, generalizing the results we introduced in Section 2.4.4. We present some elementary examples to demonstrate how to apply these results.

Throughout this and the following sections, we assume a finite time horizon $T > 0$ and a stochastic basis $\big(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t\in[0,T]}\big)$ that satisfies the usual conditions. Let $B = (B)_{t\in[0,T]}$ be a d-dimensional $(\mathcal{F}_t)_{t\in[0,T]}$ Brownian motion, where $d \geq 1$.

4.1 Motivation

To give a motivation for McKean-Vlasov stochastic differential equations, we consider an example related to physics. This is a natural choice for a motivation since, as mentioned in the introduction, the theory of MVSDEs was initiated by physics. This example is inspired by [CD18b, 2.1.2].

We want to model a system of N weakly interacting particles on some time interval $[0, T]$, where $T > 0$. For every $i = 1, 2, ..., N$ we model the position of a particle by a stochastic process $X^i = (X^i_t)_{t\in[0,T]}$. We denote by an $(B_1, B_2, ..., B_N)$ N-dimensional Brownian motion. In our model we assume that each particle solves the following stochastic different equation

$$\begin{cases} \mathrm{d}X^i_t = \sigma(t, X^i_t, \mu_N)\,\mathrm{d}B^i_t + b(t, X^i_t, \mu_N)\,\mathrm{d}t \\ X^i_0 = x^i_0, \end{cases}$$

where x^i_0 is the initial position and

$$\mu_N := \frac{1}{N}\sum_{i=1}^{N} X_i.$$

The term μ_N gives the dependence on the positions of the other particles.

To model the weak interaction, we assume that, when N is large enough, for all $t \in [0, T]$ the particles $(X^i_t)^N_{i=1}$ are behaving approximately like independent particles with identical distributions. This lets us use the *strong law of large numbers* [GG18, Proposition 8.2.6] to obtain

$$\mu_N \underset{a.s.}{\longrightarrow} \mathbb{E}X^1_t.$$

as N tends to infinity. Since we assume that the particles have identical distribution, we have that $\mathbb{E}X_t^i = \mathbb{E}X_t^1$ for all $i = 1, 2, \ldots$. This means that, if N is large enough, we may approximate individual particles with the following stochastic differential equation:

$$\begin{cases} \mathrm{d}X_t^i = \sigma(t, X_t^i, \mathbb{E}X_t^i)\,\mathrm{d}B_t^i + b(t, X_t^i, \mathbb{E}X_t^i)\,\mathrm{d}t \\ X_0^i = x_0^i. \end{cases} \tag{4.1}$$

This equation is a special case of McKean-Vlasov stochastic different equations, as we will see in the next section.

4.2 Formulation

We start by giving a formal definition for McKean-Vlasov stochastic differential equations. The definition of MVSDE is given in [BMM19, Section 2.1] and is similar to Definition 2.25.

Assume an initial value $x_0 \in \mathbb{R}^d$ and an open set $D \subseteq \mathbb{R}^d$. Assume that the coefficients

$$b : [0, T] \times D \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$$

and

$$\sigma : [0, T] \times D \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^{d\times d}$$

are Borel measurable. We consider the following type of stochastic differential equations,

$$\begin{cases} \mathrm{d}X_t = \sigma(t, X_t, \mathbb{P}_{X_t})\,\mathrm{d}B_t + b(t, X_t, \mathbb{P}_{X_t})\,\mathrm{d}t \\ X_0 = x_0 \end{cases} \tag{4.2}$$

which are called *McKean-Vlasov stochastic differential equations* or sometimes *mean-field SDEs*. An adapted and path-wise continuous process $X = (X_t)_{t\in I}$ is called a solution to (4.2) provided that

(Sol1) $X_t(\omega) \in D$ for all $t \in [0, T]$ and $\omega \in \Omega$,

(Sol2) $X_0 \equiv x_0$, and

(Sol3) we have that

$$X_t = x_0 + \int_0^t \sigma(s, X_s, \mathbb{P}_{X_s})\,\mathrm{d}B_s + \int_0^t b(s, X_s, \mathbb{P}_{X_s})\,\mathrm{d}s,$$

where

$$\int_0^t b(s, X_s, \mathbb{P}_{X_s})\,\mathrm{d}s = \left(\int_0^t b(s, X_s, \mathbb{P}_{X_s})_1\,\mathrm{d}s, \ldots, \int_0^t b(s, X_s, \mathbb{P}_{X_s})_d\,\mathrm{d}s\right),$$

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

If not mentioned otherwise, then we assume that $D = \mathbb{R}^d$.

Remark 4.1. Since we have defined the metric W_2 in the space $\mathcal{P}_2(\mathbb{R}^d)$, the Borel σ-algebra $\mathcal{B}(\mathcal{P}_2(\mathbb{R}^d)) = \mathcal{B}((\mathcal{P}_2(\mathbb{R}^d), W_2))$ is well-defined, hence the Borel-measurability of the coefficients σ and b is also well-defined: σ is Borel-measurable, if it is $(\mathcal{B}([0,T]) \otimes \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathcal{P}_2(\mathbb{R}^d)), \mathcal{B}(\mathbb{R}^{d\times d}))$-measurable, and b is Borel-measurable, if it is $(\mathcal{B}([0,T]) \otimes \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathcal{P}_2(\mathbb{R}^d)), \mathcal{B}(\mathbb{R}^d))$-measurable.

4.3 Examples

Next we consider simple examples of McKean-Vlasov stochastic differential equations. We will discuss the uniqueness of solutions in later sections. Since the presence of distribution variables makes the equations significantly more complicated and prohibits using the classical Itô's formula, we consider only the cases where the solution can be found by guessing. We also introduce two common types of how the distribution variable is used in the equations. Definitions of different interactions are mentioned in [Car16, Section 1.3.2].

Our first goal is to see an example of the interaction through marginal distributions. Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ be a function. In many applications, there exists a function $\tilde{b} : [0,T] \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ such that

$$b(t, x, \mu) = \tilde{b}(t, x, \langle \varphi, \mu \rangle),$$

where the angular bracket operator is defined by

$$\langle \varphi, \mu \rangle := \int_{\mathbb{R}^d} \varphi(u) \, \mathrm{d}\mu(u).$$

This is called *mean-field interaction of scalar type*, and it is the simplest case of interaction. If $X : \Omega \to \mathbb{R}^d$ is a random variable, then by the change of variable formula

$$\langle \varphi, \mathbb{P}_X \rangle = \int_{\mathbb{R}^d} \varphi(u) \, \mathrm{d}\mathbb{P}_X(u) = \int_{\Omega} \varphi(X(\omega)) \, \mathrm{d}\mathbb{P}(\omega) = \mathbb{E}\varphi(X).$$

If $\varphi(x) = x^n$ for some $n = 1, 2, ...$, then $\langle \varphi, \mathbb{P}_X \rangle = \mathbb{E}\,|X|^n$, that is, the n-th moment of the random variable X. Next we consider two examples of scalar type interaction.

Example 4.2. Assume that $\tilde{X} = (\tilde{X}_t)_{t\in[0,T]}$ solves the SDE

$$\begin{cases} \mathrm{d}X_t = \sigma(t, X_t) \, \mathrm{d}B_t \\ X_0 = 0. \end{cases}$$

Then $\tilde{X}_t = \int_0^t \sigma(s, \tilde{X}_s)\,\mathrm{d}B_s$. By the martingale property of stochastic integral one has

$$\mathbb{E} \int_0^t \sigma(s, \tilde{X}_s)\,\mathrm{d}B_s = 0,$$

hence $\tilde{X}$ is also a solution to the MVSDE

$$\begin{cases} \mathrm{d}X_t = \sigma(t, X_t)\,\mathrm{d}B_t + \mathbb{E}X_t\,\mathrm{d}t \\ X_0 = 0. \end{cases}$$

For example, if $\sigma \equiv 1$, then the solution is $\tilde{X}_t = \int_0^t 1\,\mathrm{d}B_s = B_t$.

If σ also has dependence on the distribution, then we can construct the MVSDE (4.1) we considered in Section 4.1.

Example 4.3. Assume Borel-measurable maps

$$\tilde{\sigma} : [0, T] \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{d \times d}$$

and

$$\tilde{b} : [0, T] \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d.$$

Define the coefficients by

$$\sigma(t, x, \mu) := \tilde{\sigma}\left(t, x, \int_{\mathbb{R}^d} u\,\mathrm{d}\mu(u)\right)$$

and

$$b(t, x, \mu) := \tilde{b}\left(t, x, \int_{\mathbb{R}^d} u\,\mathrm{d}\mu(u)\right).$$

Now the corresponding MVSDE is

$$\begin{cases} \mathrm{d}X_t = \tilde{\sigma}(t, X_t, \mathbb{E}X_t)\,\mathrm{d}B_t + \tilde{b}(t, X_t, \mathbb{E}X_t)\,\mathrm{d}t \\ X_0 = x_0, \end{cases}$$

which is what we have in Section 4.1.

We may also have more complicated types of interaction. Assume that there exists a function $\tilde{b} : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ such that

$$b(t, x, \mu) = \int_{\mathbb{R}^d} \tilde{b}(t, x, u)\,\mathrm{d}\mu(u).$$

This is called *interaction of order 1*. It can be expanded to the higher orders by

$$b(t, x, \mu) = \int_{\mathbb{R}^d} \cdots \int_{\mathbb{R}^d} \tilde{b}(t, x, u_1, ..., u_n)\,\mathrm{d}\mu(u_1) \cdots \mathrm{d}\mu(u_n).$$

In this case we have *interaction of order n*. We omit this kind of interaction in our examples.

As we will see in the following example, the solution is not always unique. Sometimes there might exists uncountably many solutions.

Example 4.4. Consider the following MVSDE

$$\begin{cases} \mathrm{d}X_t = \mathbb{1}_{(0,\infty)}(t)\sqrt{\dfrac{\mathbb{E}X_t^2}{t}}\,\mathrm{d}B_t \\ X_0 = 0. \end{cases} \tag{4.3}$$

Here

$$\sigma(t,x,\mu) = \mathbb{1}_{(0,\infty)}(t)\sqrt{\frac{1}{t}\int_{\mathbb{R}} u^2\,\mathrm{d}\mu(u)} = \begin{cases} 0, & t = 0 \\ \sqrt{\dfrac{1}{t}\displaystyle\int_{\mathbb{R}} u^2\,\mathrm{d}\mu(u)} & t > 0. \end{cases}$$

We recall that $\mathbb{E}B_t^2 = t$ for all $t \in [0,T]$, which lets us guess that the solution is $X_t = \lambda B_t$ for some constant $\lambda > 0$. Indeed now we have

$$\begin{aligned} \int_0^t \mathbb{1}_{(0,\infty)}(s)\sqrt{\frac{\mathbb{E}\left(\lambda B_s\right)^2}{s}}\,\mathrm{d}B_s &= \int_0^t \mathbb{1}_{(0,\infty)}(s)\sqrt{\frac{\lambda^2 s}{s}}\,\mathrm{d}B_s \\ &= \int_0^t \mathbb{1}_{(0,\infty)}(s)\lambda\,\mathrm{d}B_s = \lambda B_t. \end{aligned}$$

Since this holds for all $\lambda > 0$, the MVSDE (4.3) has uncountably many solutions.

In some cases there exists no solution, which we show in the next example.

Example 4.5. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a bounded Borel measurable function. Let $x_0 \in \mathbb{R}$. Consider the MVSDE

$$\begin{cases} \mathrm{d}X_t = \varphi(\mathbb{E}X_t^2)\,\mathrm{d}B_t \\ X_0 = x_0. \end{cases}$$

We want to find a process $X = (X_t)_{t\in[0,T]}$ satisfying

$$X_t = x_0 + \int_0^t \varphi(\mathbb{E}X_s^2)\,\mathrm{d}B_s$$

for $t \in [0,T]$ almost surely. First we note that

$$\begin{aligned}
\mathbb{E}X_t^2 &= \mathbb{E}\left(x_0 + \int_0^t \varphi(\mathbb{E}X_s^2)\,\mathrm{d}B_s\right)^2 \\
&= x_0^2 + 2x_0\mathbb{E}\int_0^t \varphi(\mathbb{E}X_s^2)\,\mathrm{d}B_s + \mathbb{E}\left[\int_0^t \varphi(\mathbb{E}X_s^2)\,\mathrm{d}B_s\right]^2 \\
&= x_0^2 + \mathbb{E}\left[\int_0^t \varphi(\mathbb{E}X_s^2)\,\mathrm{d}B_s\right]^2.
\end{aligned}$$

Itô's isometry 2.21 yields

$$\begin{aligned}
\mathbb{E}\left[\int_0^t \varphi(\mathbb{E}X_s^2)\,\mathrm{d}B_s\right]^2 &= \mathbb{E}\int_0^t \varphi(\mathbb{E}X_s^2)^2\,\mathrm{d}s \\
&= \int_0^t \varphi(\mathbb{E}X_s^2)^2\,\mathrm{d}s.
\end{aligned}$$

The term $\varphi(\mathbb{E}X_s^2)^2$ does not depend on $\omega \in \Omega$, which implies the last equality.

We let $\psi(x) := \varphi(x)^2$ and $f(t) := \mathbb{E}X_t^2$. We can write the previous equation as an integral equation

$$\begin{aligned}
f(t) = \mathbb{E}X_t^2 &= x_0^2 + \int_0^t \varphi(\mathbb{E}X_s^2)^2\,\mathrm{d}s \\
&= x_0^2 + \int_0^t \psi(f(s))\,\mathrm{d}s.
\end{aligned}$$

To show that (4.5) has no solution, it is sufficient to show that this integral equation has no solution.

For example, let $x_0 = 0$ and $\psi(x) := \mathbb{1}_{\mathbb{Q}}(x)$. Now the integral equation looks like

$$f(t) = \int_0^t \mathbb{1}_{\mathbb{Q}}(f(s))\,\mathrm{d}s, \quad t \in [0,T].$$

We show that this equation does not have a solution. We consider three different cases.

(1) If $f \equiv c$ for some constant $c \in \mathbb{R}$, then it has to be $c = 0$, because $f(0) = 0$. However, now we have

$$\int_0^t \mathbb{1}_{\mathbb{Q}}(0)\,\mathrm{d}s = t,$$

which is a contradiction.

(2) If f is strictly increasing, then it has to hold that

$$\lambda(\{t \in [0,T] \mid f(t) \in \mathbb{Q}\}) = 0,$$

which implies $\mathbb{1}_{\mathbb{Q}}(f) = 0$ almost everywhere. It follows from the fact that $f^{-1}(\mathbb{Q})$ is at most countable in this case. This implies that

$$\int_0^t \mathbb{1}_{\mathbb{Q}}(f(s))\,\mathrm{d}s = 0,$$

which is a contradiction.

(3) Assume that there exists an interval $[a,b] \subset (0,T]$, $b > a$, such that $f(t) = q > 0$ for all $t \in [a,b]$. If $q \notin \mathbb{Q}$, then

$$f(t) = \int_a^{\min\{b,t\}} \mathbb{1}_{\mathbb{Q}}(f(s))\,\mathrm{d}s = \int_a^{\min\{b,t\}} \mathbb{1}_{\mathbb{Q}}(q)\,\mathrm{d}s = 0,$$

thus it has to hold that $q \in \mathbb{Q}$. However, now

$$f(t) = \int_a^{\min\{b,t\}} \mathbb{1}_{\mathbb{Q}}(f(s))\,\mathrm{d}s = \int_a^{\min\{b,t\}} \mathbb{1}_{\mathbb{Q}}(q)\,\mathrm{d}s = \min\{b,t\} - a,$$

which is not a constant.

This implies that the MVSDE

$$\begin{cases} \mathrm{d}X_t = \sqrt{\mathbb{1}_{\mathbb{Q}}(\mathbb{E}X_t^2)}\,\mathrm{d}B_t = \mathbb{1}_{\mathbb{Q}}(\mathbb{E}X_t^2)\,\mathrm{d}B_t \\ X_0 = 0 \end{cases}$$

has no solution.

4.4 Existence and uniqueness of a solution

In this section our goal is to formulate similar existence and uniqueness results as we did for ordinary SDEs in Section 2.4.4. The first result generalizes the theorem that states the existence and uniqueness of a solution under the Lipschitz continuity and linear growth condition to MVSDEs. We formulate this theorem and consider a simple example, but we do not prove this theorem. Our second theorem is a generalization of the theorem of Yamada-Tanaka, mentioned in Theorem 2.28, but we generalize the setting to a specific multidimensional case.

4.4.1 Existence and uniqueness under Lipschitz condition

In Theorem 2.27 we saw that if the coefficient functions are globally Lipschitz continuous and satisfy the linear growth condition, then there exists a unique solution. As $\mathcal{P}_2(\mathbb{R}^d)$ equipped with the Wasserstein distance W_2 is a metric space, we may formulate a similar theorem for MVSDEs. We give the same formulation as in [BMM19, Section 3.1].

We assume that the coefficients of the MVSDE (4.2) satisfy the following properties:

(L1) There exists a constant $K > 0$ such that

$$\|b(t, x, \mu)\| \leq K (1 + \|x\|)$$

and

$$\|\sigma(t, x, \mu)\| \leq K (1 + \|x\|)$$

for all $(t, x, \mu) \in [0, T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$.

(L2) There exists a constant $L > 0$ such that

$$\|b(t, x, \mu) - b(t, y, \nu)\| \leq L [\|x - y\| + W_2(\mu, \nu)]$$

and

$$\|\sigma(t, x, \mu) - \sigma(t, y, \nu)\| \leq L [\|x - y\| + W_2(\mu, \nu)]$$

for all $t \in [0, T]$, $x, y \in \mathbb{R}^d$ and $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$.

Theorem 4.6. *Under assumptions (L1) and (L2), the MVSDE* (4.2) *has a unique solution. Moreover, if $X = (X_t)_{t\in[0,T]}$ is the solution, it holds that*

$$\mathbb{E}\left[\sup_{t\in[0,T]} \|X_t\|^2\right] < \infty.$$

A complete proof can be found in [CD18a, Theorem 4.21]. The proof is given for a slightly different setting, where the linear growth condition (L1) is replaced with another condition. However, the proof in our setting is identical, since the corresponding condition in [CD18a] is only required to conclude that a unique solution exists for an ordinary SDE.

We demonstrate this theorem in the following example.

Example 4.7. We want to show that there exists a unique solution to the following MVSDE

$$\begin{cases} \mathrm{d}X_t = \cos(X_t)\,\mathrm{d}B_t + \mathbb{E}\sin(X_t)\,\mathrm{d}t, \\ X_0 = x_0. \end{cases}$$

We recall that sine and cosine functions are bounded 1-Lipschitz functions. Hence the coefficient σ clearly satisfies (L1) and (L2), so we only need to check that the coefficient b is Lipschitz continuous.

We have that

$$b(t,x,\mu) = \int_{\mathbb{R}} \sin(u)\,\mathrm{d}\mu(u).$$

Now, using Kantovich-Rubinstein duality 3.5 we obtain

$$\begin{aligned}
|b(t,x,\mu) - b(t,y,\nu)| &= \left| \int_{\mathbb{R}} \sin(u)\,\mathrm{d}\mu(u) - \int_{\mathbb{R}} \sin(t)\,\mathrm{d}\nu(t) \right| \\
&= \left| \int_{\mathbb{R}} \sin(u)\,\mathrm{d}(\mu-\nu)(u) \right| \\
&\leq \sup \left\{ \left| \int_{\mathbb{R}^d} h(u)\,\mathrm{d}(\mu-\nu)(u) \right| \;\Big|\; h \in \mathrm{Lips}_1(\mathbb{R}) \right\} \\
&= W_1(\mu,\nu) \\
&\leq |x-y| + W_1(\mu,\nu) \\
&\leq |x-y| + W_2(\mu,\nu),
\end{aligned}$$

where we use Lemma 3.7 to get the final inequality. This implies that there exists a unique solution to (4.7).

4.4.2 Generalization of Yamada-Tanaka theorem

Next we want to generalize the uniqueness theorem of Yamada and Tanaka, introduced in Theorem 2.28. We consider the case where only the coefficient b depends on the distribution variable. In the one dimensional case, this result is proven in [BMM19, Section 3.2]. However, by adapting this proof, we may generalize the theorem even further and consider a specific multidimensional case.

$$\sigma_i : [0,T] \times \mathbb{R} \to \mathbb{R}$$

and

$$b_i : [0,T] \times \mathbb{R} \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$$

be bounded Borel measurable functions. We define the coefficients $\sigma : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ and $b : [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ such that

$$b(t,(x_1,...,x_d),\mu) := (b_1(t,x_1,\mu),...,b_d(t,x_d,\mu))$$

and

$$\sigma(t,(x_1,...,x_d)) := \mathrm{Diag}(\sigma_1(t,x_1),...,\sigma_d(t,x_d)),$$

where Diag denotes a $d \times d$ diagonal matrix, that is

$$\mathrm{Diag}(\sigma_1(t,x_1),...,\sigma_d(t,x_d)) = \begin{bmatrix} \sigma_1(t,x_1) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_d(t,x_d) \end{bmatrix}.$$

Assume that $X = (X^1, ..., X^d)$ solves

$$\begin{cases} \mathrm{d}X_t = \sigma(t,X_t)\,\mathrm{d}B_t + b(t,X_t,\mathbb{P}_{X_t})\,\mathrm{d}t \\ X_0 = x_0 = (x_0^1,...,x_0^d). \end{cases}$$

By definition this is equivalent to the system of one-dimensional MVSDEs where for each $i = 1, ..., d$, X^i is a solution to

$$\begin{cases} \mathrm{d}X_t^i = \sigma_i(t,X_t^i)\,\mathrm{d}B_t^i + b_i(t,X_t^i,\mathbb{P}_{X_t})\,\mathrm{d}t \\ X_0^i = x_0^i. \end{cases} \tag{4.4}$$

It should be noted that here the coefficient for the drift term depends on the law of the whole d-dimensional process, not just its ith component.

In this setting, we may give sufficient conditions for the uniqueness of a solution. We assume that for all $i = 1, ..., d$ the following conditions are satisfied:

(A1) The function b_i is Lipschitz-continuous with respect to the distribution variable, that is, there exists a constant $C > 0$ such that

$$|b_i(t,x,\mu) - b_i(t,x,\nu)| \leq CW_1(\mu,\nu)$$

for all $x \in \mathbb{R}$, $t \in [0,T]$ and $(\mu,\nu) \in \mathcal{P}_1(\mathbb{R}^d) \otimes \mathcal{P}_1(\mathbb{R}^d)$

(A2) There exists a strictly increasing function $\rho : [0,\infty) \to [0,\infty)$ satisfying $\rho(0) = 0$ and

$$\int_0^\epsilon \frac{1}{\rho^2(u)}\,\mathrm{d}u = \infty$$

for every $\epsilon > 0$, and $|\sigma_i(t,x) - \sigma_i(t,y)| \leq \rho(|x-y|)$ for all $t \in [0,T]$ and $x, y \in \mathbb{R}$.

(A3) There exists a strictly increasing concave function $\kappa : [0,\infty) \to [0,\infty)$ satisfying $\kappa(0) = 0$ and

$$\int_0^\epsilon \frac{1}{\kappa(u)}\,\mathrm{d}u = \infty$$

for every $\epsilon > 0$, and $|b_i(t,x,\mu) - b_i(t,y,\mu)| \leq \kappa(|x-y|)$ for all $t \in [0,T]$, $x, y \in \mathbb{R}$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.

With these conditions we can formulate the following theorem that gives us the uniqueness of a solution.

Theorem 4.8. *Under conditions (A1), (A2) and (A3), a solution to* (4.4.2) *is unique.*

Before we can prove this theorem, we need two lemmas. The first lemma is known as *Bihari-LaSalle* inequality, which is a non-linear generalization of Gronwall's inequality.

Lemma 4.9 (Bihari–LaSalle, [Mao07, Section 1.8, theorem 8.2])**.** *Assume constants $T > 0$ and $c > 0$. Let $f, u : [0,T] \to [0,\infty)$ be continuous functions. Let $\kappa : [0,\infty) \to [0,\infty)$ be a continuous and increasing function such that $\kappa(x) > 0$ for all $x > 0$. If the function u is bounded and satisfies the following inequality*

$$u(t) \leq c + \int_0^t f(s)\kappa(u(s))\,\mathrm{d}s$$

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

$$u(t) \leq G^{-1}\left(G(c) + \int_0^t f(s)\,\mathrm{d}s\right)$$

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

$$G(c) + \int_0^t f(s)\,\mathrm{d}s \in Dom(G^{-1}), \tag{4.5}$$

where

$$G(x) := \int_1^x \frac{1}{\kappa(u)}\,\mathrm{d}u,$$

for $x > 0$.

Proof. Let $v(t) := c + \int_0^t f(s)\kappa(u(s))\,\mathrm{d}s$. We differentiate v to get

$$v'(t) = f(t)\kappa(u(t)).$$

By the chain rule we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}G(v(t)) = v'(t)G'(v(t)) = \frac{v'(t)}{\kappa(v(t))} = \frac{f(t)\kappa(u(t))}{\kappa(v(t))}.$$

Integrating from 0 to t yields

$$\begin{aligned}\int_0^t \left[\frac{\mathrm{d}}{\mathrm{d}s}G(v(s))\right]\mathrm{d}s &= G(v(t)) - G(v(0)) \\ &= G(v(t)) - G(c) = \int_0^t \frac{f(s)\kappa(u(s))}{\kappa(v(s))}\,\mathrm{d}s.\end{aligned}$$

Since κ is a strictly increasing function we may apply the estimate $\kappa(u(s)) \leq \kappa(v(s))$ to see that

$$G(v(t)) - G(c) = \int_0^t \frac{f(s)\kappa(u(s))}{\kappa(v(s))}\,\mathrm{d}s \leq \int_0^t \frac{f(s)\kappa(v(s))}{\kappa(v(s))}\,\mathrm{d}s = \int_0^t f(s)\,\mathrm{d}s.$$

Therefore,

$$G^{-1}(G(v(t))) = v(t) \leq G^{-1}\left(G(c) + \int_0^t f(s)\,\mathrm{d}s\right).$$

for $t \in [0, T]$ where $T > 0$ is chosen so that it satisfies the property (4.5). This completes the proof since $u(t) \leq v(t)$ for all $t \geq 0$, □

With Bihari–LaSalle inequality we can easily prove Gronwall's inequality.

Lemma 4.10 (Gronwall's inequality, [Mao07, Section 1.8, theorem 8.1])**.** *Let $A, B, T \geq 0$ and let $u : [0, T] \to [0, \infty)$ be a continuous function satisfying*

$$u(t) \leq A + B\int_0^t u(s)\,\mathrm{d}s$$

for all $u \in [0, T]$. Then $u(t) \leq Ae^{Bt}$ for all $t \in [0, T]$.

Proof. By choosing $\kappa(x) := x$ and $f := B$ one has

$$G(x) = \int_1^x \frac{1}{u}\,\mathrm{d}u = \log(x) - \log(1) = \log(x)$$

and $G^{-1}(x) = e^x$. Then, by Lemma 4.9 we have

$$u(t) \leq \exp(\log(A) + \int_0^t B\,\mathrm{d}s) = Ae^{Bt}$$

for $t \in [0, T]$, in the case $A > 0$. If $A = 0$, then we use the estimate

$$u(t) \leq B\int_0^t u(s)\,\mathrm{d}s < \epsilon + B\int_0^t u(s)\,\mathrm{d}s$$

for every $\epsilon > 0$. Now

$$u(t) \leq \exp\left(\log(\epsilon) + \int_0^t B\,\mathrm{d}s\right) \to 0$$

as ϵ tends to 0.

□

Now we can prove Theorem 4.8.

Proof of Theorem 4.8. We define the norm

$$\|x\|_{1,d} := \sum_{i=1}^{d} |x_i|$$

in $\mathbb{R}^d$. We see that

$$\|x\|_{2,d} := \|x\| = \sqrt{\sum_{k=1}^{d} |x_k|^2} \leq \sum_{k=1}^{d} \sqrt{|x_k|}^2 = \sum_{k=1}^{d} |x_k| = \|x\|_{1,d}$$

and by the Cauchy-Schwartz inequality we have that

$$\|x\|_{1,d}^2 = \left(\sum_{k=1}^{d} |x_k|\right)^2 \leq d \sum_{k=1}^{d} |x_k|^2 = \|x\|_{2,d}^2 .$$

Hence

$$\|x\|_{2,d} \leq \|x\|_{1,d} \leq \sqrt{d}\, \|x\|_{1,d} .$$

This implies that the norms $\|\cdot\|_{1,d}$ and $\|\cdot\|_{2,d}$ are equivalent.

We assume two processes $X = (X^1, ..., X^d)$ and $Y = (Y^1, ..., Y^d)$ that solve (4.4.2). Our goal is to show that

$$\mathbb{E}\, \|X_t - Y_t\|_{1,d} = 0$$

for all $t \in [0, T]$, which implies that X and Y are indistinguishable.

By assumption (A2) we have that

$$\int_0^{\epsilon} \frac{1}{\rho(u)^2}\, \mathrm{d}u = \infty,$$

for all $\epsilon > 0$. It follows that for every $\xi > 0$ there exists $a \in (0, 1)$ such that

$$\int_a^1 \frac{1}{\rho(u)^2}\, \mathrm{d}u = \xi.$$

This lets us construct a sequence $(a_n)_{n=1}^{\infty}$ of real numbers such that

$$1 > a_1 > a_2 > ... > a_n > a_{n-1} > ... > 0$$

and

$$\int_{a_1}^1 \frac{1}{\rho(u)^2}\, \mathrm{d}u = 1 \quad \text{and} \quad \int_{a_n}^{a_{n-1}} \frac{1}{\rho(u)^2}\, \mathrm{d}u = n$$

for all $n \geq 2$. Moreover, we see that $a_n \to 0$ as $n \to \infty$.

Next we construct a sequence of functions $(\psi_n)_{n=1}^{\infty}$, $\psi_n : \mathbb{R} \to \mathbb{R}$, such that for every $n \in \mathbb{N}$ we have that:

(1) ψ_n is continuous,

(2) we have
$$\{x \in \mathbb{R} \mid \psi_n(x) \neq 0\} \subseteq (a_n, a_{n-1}),$$

(3) and for all $x \in \mathbb{R}$ one has
$$0 \leq \psi_n(x) \leq \frac{2}{n\rho(x)^2}$$
and
$$\int_{a_n}^{a_{n-1}} \psi_n(u)\,\mathrm{d}u = 1.$$

The idea in this construction is that for all $n \in \mathbb{N}$ we approximate the function
$$x \mapsto \mathbb{1}_{(a_n,a_{n-1})}(x)\frac{1}{n\rho(x)^2}$$
with a continuous function such that the integral over $\mathbb{R}$ is the same. We do not go in details why a function like this exists for every $n \in \mathbb{N}$.

For $n \in \mathbb{N}$ we let
$$\varphi_n(x) := \int_0^{|x|} \int_0^y \psi_n(u)\,\mathrm{d}u\,\mathrm{d}y.$$

It should be noted that now we think the integral as a Riemann integral to get the required properties. Clearly $\varphi_n \in \mathcal{C}^2(\mathbb{R})$. We see that
$$\varphi_n'(x) = \int_0^x \psi_n(u)\,\mathrm{d}u$$
for $x \geq 0$, and $\varphi_n'(x) = -\varphi_n'(-x)$ for $x < 0$. Therefore
$$|\varphi_n'(x)| = \int_0^{|x|} \psi_n(u)\,\mathrm{d}u \leq \int_{a_n}^{a_{n-1}} \psi_n(u)\,\mathrm{d}u = 1$$
for all $n \geq 1$. Since a_n converges to 0 as $n \to \infty$, it follows that
$$\int_0^y \psi_n(u)\,\mathrm{d}u \to 1$$
as n tends to ∞ for all $y > 0$. Therefore the sequence $(\varphi_n)_{n=1}^{\infty}$ converges to the function $\varphi(x) := |x|$.

Next we fix $i = 1, ..., d$ and consider processes X^i and Y^i that are solutions to (4.4). Let $Z = X^i - Y^i$. We see that Z is an Itô process with representation

$$Z_t = X_t^i - Y_t^i = 0 + \int_0^t \left(\sigma_i(s, X_s^i) - \sigma_i(s, Y_s^i)\right) \mathrm{d}B_s^i + \int_0^t \left(b_i(s, X_s^i, \mathbb{P}_{X_s}) - b_i(s, Y_s^i, \mathbb{P}_{Y_s})\right) \mathrm{d}s.$$

We apply Theorem 2.24 to the process Z and function φ_n to get that

$$\begin{aligned}\varphi_n(Z_t) = 0 &+ \int_0^t \varphi_n'(Z_s) \left(\sigma_i(s, X_s^i) - \sigma_i(s, Y_s^i)\right) \mathrm{d}B_s^i \\ &+ \int_0^t \varphi_n'(Z_s) \left(b_i(s, X_s^i, \mathbb{P}_{X_s}) - b_i(s, Y_s^i, \mathbb{P}_{Y_s})\right) \mathrm{d}s \\ &+ \frac{1}{2} \int_0^t \varphi_n''(Z_s) \left(\sigma_i(s, X_s^i) - \sigma_i(s, Y_s^i)\right)^2 \mathrm{d}s.\end{aligned}$$

It should be noted that since φ_n' and σ_i are bounded and measurable, the stochastic integral exists and

$$\mathbb{E} \int_0^t \varphi_n'(Z_s) \left(\sigma_i(s, X_s^i) - \sigma_i(s, Y_s^i)\right) \mathrm{d}B_s^i = 0.$$

Therefore

$$\begin{aligned}\mathbb{E}\left[\varphi_n(Z_t)\right] &= \mathbb{E}\left[\int_0^t \varphi_n'(Z_s) \left(b_i(s, X_s^i, \mathbb{P}_{X_s}) - b_i(s, Y_s^i, \mathbb{P}_{Y_s})\right) \mathrm{d}s\right] \\ &\quad + \frac{1}{2}\mathbb{E}\left[\int_0^t \varphi_n''(Z_s) \left(\sigma_i(s, X_s^i) - \sigma_i(s, Y_s^i)\right)^2 \mathrm{d}s\right] \\ &= I_1 + I_2,\end{aligned}$$

where I_1 is the first term on the right-hand side and I_2 the second term.

First, we estimate I_2 by

$$\begin{aligned}|I_2| &= \frac{1}{2}\left|\mathbb{E}\left[\int_0^t \varphi_n''(X_s^i - Y_s^i) \left(\sigma_i(s, X_s^i) - \sigma_i(s, Y_s^i)\right)^2 \mathrm{d}s\right]\right| \\ &\leq \frac{1}{2}\left|\mathbb{E}\left[\int_0^t \psi_n(X_s^i - Y_s^i)\rho\left(X_s^i - Y_s^i\right)^2 \mathrm{d}s\right]\right| \\ &\leq \frac{1}{2}\left|\mathbb{E}\left[\int_0^t \frac{2}{n\rho(X_s^i - Y_s^i)^2}\rho\left(X_s^i - Y_s^i\right)^2 \mathrm{d}s\right]\right| \\ &= \frac{t}{n}.\end{aligned}$$

Clearly $\frac{t}{n} \to 0$ as $n \to \infty$.

We continue by estimating the term I_1. We have shown that φ' is bounded by 1. Hence

$$\begin{aligned}&\left|\varphi_n'(X_t^i - Y_t^i)\left(b_i(t, X_t^i, \mathbb{P}_{X_t}) - b_i(t, Y_t^i, \mathbb{P}_{Y_t})\right)\right| \\ &\le \left|\left(b_i(t, X_t^i, \mathbb{P}_{X_t}) - b_i(t, Y_t^i, \mathbb{P}_{Y_t})\right)\right|.\end{aligned}$$

We apply triangle inequality to obtain

$$\begin{aligned}\left|\left(b_i(t, X_t^i, \mathbb{P}_{X_t}) - b_i(t, Y_t^i, \mathbb{P}_{Y_t})\right)\right| &= \big|\left[(b_i(t, X_t^i, \mathbb{P}_{Y_t}) - b_i(t, Y_t^i, \mathbb{P}_{Y_t})\right] \\ &\quad - \left[b_i(t, X_t^i, \mathbb{P}_{Y_t}) - b_i(t, X_t^i, \mathbb{P}_{X_t})\right]\big| \\ &\le \left|b_i(t, X_t^i, \mathbb{P}_{Y_t}) - b_i(t, Y_t^i, \mathbb{P}_{Y_t})\right| \\ &\quad + \left|b_i(t, X_t^i, \mathbb{P}_{Y_t}) - b_i(t, X_t^i, \mathbb{P}_{X_t})\right|.\end{aligned}$$

The Lipschitz property (A1) implies

$$\begin{aligned}\left|b_i(t, X_t^i, \mathbb{P}_{Y_t}) - b_i(t, X_t^i, \mathbb{P}_{X_t})\right| &\le C W_1(\mathbb{P}_{Y_t}, \mathbb{P}_{X_t}) \\ &\le C\mathbb{E}\left\|X_t - Y_t\right\|,\end{aligned}$$

where Lemma 3.4 implies the latter inequality. Then

$$C\mathbb{E}\left\|X_t - Y_t\right\| \le C\mathbb{E}\left\|X_t - Y_t\right\|_{1,d}.$$

Finally, assumption (A3) gives us

$$\left|b_i(t, X_t^i, \mathbb{P}_{Y_t}) - b_i(t, Y_t^i, \mathbb{P}_{Y_t})\right| \le \kappa\left(\left|X_t^i - Y_t^i\right|\right).$$

Now we have that

$$\begin{aligned}|I_1| &\le \mathbb{E}\left[\int_0^t \left|\varphi_n'(Z)\left(b_i(s, X_s^i, \mathbb{P}_{X_s}) - b_i(s, Y_s^i, \mathbb{P}_{Y_s})\right)\right| \mathrm{d}s\right] \\ &\le \mathbb{E}\left[\int_0^t \left(\kappa\left(\left|X_s^i - Y_s^i\right|\right) + C\mathbb{E}\left\|X_s - Y_s\right\|_{1,d}\right) \mathrm{d}s\right] \\ &= \mathbb{E}\int_0^t \kappa\left(\left|X_s^i - Y_s^i\right|\right) \mathrm{d}s + C\int_0^t \mathbb{E}\left\|X_s - Y_s\right\|_{1,d} \mathrm{d}s.\end{aligned}$$

We recall that $|I_2|$ converges to 0 and the function φ_n converges to the function $\varphi(x) = |x|$ as $n \to \infty$, so letting n tend to ∞ we get the inequality

$$\mathbb{E}\left|X_t^i - Y_t^i\right| \le \mathbb{E}\int_0^t \kappa\left(\left|X_s^i - Y_s^i\right|\right) \mathrm{d}s + C\int_0^t \mathbb{E}\left\|X_s - Y_s\right\|_{1,d} \mathrm{d}s$$

for all $i = 1, ..., d$. Then

$$\begin{aligned}
\sum_{i=1}^{d} \mathbb{E}\left|X_t^i - Y_t^i\right| &= \mathbb{E}\left\|X_t - Y_t\right\|_{1,d} \\
&\leq \sum_{i=1}^{d} \mathbb{E}\int_0^t \kappa\left(\left|X_s^i - Y_s^i\right|\right) \mathrm{d}s + C\sum_{i=1}^{d}\int_0^t \mathbb{E}\left\|X_s - Y_s\right\|_{1,d} \mathrm{d}s \\
&= \sum_{i=1}^{d} \mathbb{E}\int_0^t \kappa\left(\left|X_s^i - Y_s^i\right|\right) \mathrm{d}s + Cd\int_0^t \mathbb{E}\left\|X_s - Y_s\right\|_{1,d} \mathrm{d}s.
\end{aligned}$$

Next we fix $r \in (0, T]$. Let $B := Cd$ and let

$$A(t) := \sum_{i=1}^{d} \mathbb{E}\int_0^t \kappa(\left|X_s^i - Y_s^i\right|) \mathrm{d}s$$

for $t \in [0, r]$. Define a function

$$f(t) := \mathbb{E}\left\|X_t - Y_t\right\|_{1,d}$$

for $t \in [0, t]$. Now we may write

$$f(t) \leq A(t) + B\int_0^t f(s) \mathrm{d}s \leq A(r) + B\int_0^t f(s) \mathrm{d}s$$

for all $t \in [0, r]$. Since r is fixed, we can consider $A(r)$ a constant, so we may apply Lemma 4.10 to obtain

$$\begin{aligned}
f(t) &\leq A(r)\exp(Bt) \leq A(r)\exp(BT) \\
&= M\sum_{i=1}^{d} \mathbb{E}\int_0^r \kappa(\left|X_s^i - Y_s^i\right|) \mathrm{d}s,
\end{aligned}$$

for all $t \in [0, T]$, where $M := \exp(BT)$. Since this holds for all $r \in (0, T]$, we have that

$$f(t) = \mathbb{E}\left\|X_t - Y_t\right\|_{1,d} \leq M\sum_{i=1}^{d} \mathbb{E}\int_0^t \kappa(\left|X_s^i - Y_s^i\right|) \mathrm{d}s$$

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

By Theorem 2.8 we can take the expectation inside the integral, that is

$$M\sum_{i=1}^{d} \mathbb{E}\int_0^t \kappa(\left|X_s^i - Y_s^i\right|) \mathrm{d}s = M\sum_{i=1}^{d} \int_0^t \mathbb{E}\kappa(\left|X_s^i - Y_s^i\right|) \mathrm{d}s.$$

We apply Proposition 2.6 to obtain

$$\mathbb{E}\kappa(\left|X_t^i - Y_t^i\right|) \leq \kappa(\mathbb{E}\left|X_t^i - Y_t^i\right|),$$

and therefore

$$\begin{aligned} M\sum_{i=1}^{d}\int_0^t \mathbb{E}\kappa(\left|X_s^i - Y_s^i\right|)\,\mathrm{d}s &\leq M\sum_{i=1}^{d}\int_0^t \kappa(\mathbb{E}\left|X_s^i - Y_s^i\right|)\,\mathrm{d}s \\ &= M\int_0^t\sum_{i=1}^{d}\kappa(\mathbb{E}\left|X_s^i - Y_s^i\right|)\,\mathrm{d}s. \end{aligned}$$

Since κ is an increasing function, we can continue our estimate

$$\begin{aligned} \mathbb{E}\left\|X_t - Y_t\right\|_{1,d} &\leq M\int_0^t\sum_{i=1}^{d}\kappa\left(\mathbb{E}\left|X_s^i - Y_s^i\right|\right)\mathrm{d}s \\ &\leq M\int_0^t\sum_{j=1}^{d}\kappa\left(\sum_{i=1}^{d}\mathbb{E}\left|X_s^i - Y_s^i\right|\right)\mathrm{d}s \\ &= Md\int_0^t\kappa(\mathbb{E}\left\|X_s - Y_s\right\|_{1,d})\,\mathrm{d}s. \end{aligned}$$

Next we use Lemma 4.9. For all $\epsilon \in (0,1)$ we have

$$\mathbb{E}\left\|X_t - Y_t\right\|_{1,d} \leq Md\int_0^t\kappa(\mathbb{E}\left\|X_s - Y_s\right\|_{1,d})\,\mathrm{d}s + \epsilon,$$

so by Lemma 4.9 we obtain

$$\begin{aligned} \mathbb{E}\left\|X_t - Y_t\right\|_{1,d} &\leq G^{-1}\left(\int_1^\epsilon \frac{1}{\kappa(u)}\,\mathrm{d}u + (Md)t\right) \\ &= G^{-1}\left(-\int_\epsilon^1 \frac{1}{\kappa(u)}\,\mathrm{d}u + (Md)t\right), \end{aligned}$$

where $G : (0,1) \to (-\infty, 0]$,

$$G(x) := \int_1^x \frac{1}{\kappa(u)}\,\mathrm{d}u,$$

is a bijection. Here we assume that ϵ is small enough so that

$$\int_\epsilon^1 \frac{1}{\kappa(u)}\,\mathrm{d}u > (Md)t.$$

Since G is a strictly increasing function such that $G(x) \to -\infty$ as x tends to 0 from the right-hand side, the inverse function G^{-1} is also strictly increasing with $G^{-1}(x) \to 0$ as $x \to -\infty$. It follows that

$$G^{-1}\left(-\int_{\epsilon}^{1} \frac{1}{\kappa(u)}\,\mathrm{d}u + (Md)t\right) \to 0$$

as $\epsilon \to 0$. Hence $\mathbb{E}\,\|X_t - Y_t\|_{1,d} = 0$ for all $t \in [0,T]$.

This property implies that $\|X_t(\omega) - Y_t(\omega)\|_{1,d} = 0$ almost everywhere. By the properties of a norm we have that $X_t(\omega) = Y_t(\omega)$ almost everywhere, that is,

$$\mathbb{P}\left(X_t = Y_t\right) = 1$$

for all $t \in [0,T]$. By definition X and Y are modifications of each other. Since all the trajectories of processes X and Y are continuous, using Proposition 2.17 we conclude that X and Y are indistinguishable, which completes our proof. □

Remark 4.11. It should be noted that Theorem 4.8 only implies uniqueness of the solution, but does not imply the existence. In the case we already know some solution for an MVSDE, we may apply the theorem to confirm that the solution is indeed unique.

Next we give a simple example how one can use Theorem 4.8 to prove the uniqueness of a solution. Since the theorem does not imply the existence, we have to find some solution first.

Example 4.12. We consider the following MVSDE:

$$\begin{cases} \mathrm{d}X_t = \min\left\{\sqrt{|X_t|}, 1\right\}\mathrm{d}B_t + \dfrac{\mathbb{E}X_t}{1+(\mathbb{E}X_t)^2}\,\mathrm{d}t \\ X_0 = 0. \end{cases}$$

We see that the process $X \equiv 0$ solves the equation. We apply Theorem 4.8 to prove that this solution is actually the only one. In this example the coefficients are the following:

$$\sigma(t,x) = \min\left\{\sqrt{|x|}, 1\right\}$$

and

$$b(t,x,\mu) = \varphi\left(\int_{\mathbb{R}} u\,\mathrm{d}\mu(u)\right),$$

where

$$\varphi(x) := \frac{x}{1+x^2}.$$

We check the conditions (A1), (A2) and (A3) separately, starting from the first one.

We see that

$$\frac{\mathrm{d}}{\mathrm{d}x}\varphi(x) = \frac{1-x^2}{(x^2+1)^2}$$

is bounded, which implies that φ is a Lipschitz-continuous function with some constant $L > 0$. For for all $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ we have that

$$\begin{aligned}
|b(t,x,\mu) - b(t,x,\nu)| &= \left|\varphi\left(\int_{\mathbb{R}} u \,\mathrm{d}\mu(u)\right) - \varphi\left(\int_{\mathbb{R}} u \,\mathrm{d}\nu(u)\right)\right| \\
&\leq L \left|\int_{\mathbb{R}} u \,\mathrm{d}(\mu-\nu)(u)\right| \\
&\leq L \sup\left\{\int_{\mathbb{R}} h(u)\,\mathrm{d}(\mu-\nu)(u) \mid h \in \mathrm{Lips}_1(\mathbb{R})\right\} \\
&= LW_1(\mu,\nu),
\end{aligned}$$

where the final equality follows from Theorem 3.5.

Next we verify the condition (A2). We may choose $\rho(x) := \sqrt{x}$. It is defined on $[0,\infty)$, increasing and has $\rho(0) = 0$. Furthermore,

$$\int_0^{\epsilon} \frac{1}{\rho(u)}\,\mathrm{d}u = \infty$$

for all $\epsilon > 0$. Using the properties of square root we obtain

$$\begin{aligned}
|\sigma(t,x) - \sigma(t,y)| &= \left|\min\left\{\sqrt{|x|}, 1\right\} - \min\left\{\sqrt{|y|}, 1\right\}\right| \\
&\leq \left|\sqrt{|x|} - \sqrt{|y|}\right| \\
&\leq \sqrt{|x-y|} = \rho(|x-y|)
\end{aligned}$$

for all $x, y \in \mathbb{R}$.

The coefficient b depends only on the distribution variable, hence we do not need to check the third condition (A3). Since all the conditions are satisfied, the uniqueness of the solution follows from Theorem 4.8.

5 Stability and approximation of MVSDEs

In this section we consider various stability and approximation results. In our first result we introduce an iterative method for the approximation of a solution, and we prove that under certain conditions a sequence of iterated processes converges to the unique solution of (4.2). In the next three results we consider the stability of the solution. We consider stability from the following points of views:

1. *Stability with respect to the initial condition.* We prove that if we define a map that maps the initial value to the solution of (4.2), under certain conditions this map is continuous.

2. *Stability with respect to the coefficients.* One way to approximate the solution is to define sequences of functions that converge to the coefficients. We prove that under certain assumptions solutions obtained this way eventually converge to the unique solution of (4.2).

3. *Stability with respect to the driving process.* So far we have considered MVSDEs with respect to the Brownian motion. In our final stability result we change this settting. Under sufficient conditions we may approximate the driving process with possibly simpler processes, and the solutions obtained in this way converge to the unique solution of (4.2).

5.1 Picard approximation

We start with the *Picard approximation*, which gives us a method to construct a sequence of processes that eventually converges to the unique solution of an MVSDE. This is a useful method in numeric computations.

Assume a sequence of processes $((X_t^n)_{t\in[0,T]})_{n=0}^{\infty}$ such that $X^0 \equiv x_0$. For $n \geq 0$ define a process X^{n+1} by

$$(5.1) \qquad \begin{cases} \mathrm{d}X_t^{n+1} = \sigma(t, X_t^n, \mathbb{P}_{X_t^n})\,\mathrm{d}B_t + b(t, X_t^n, \mathbb{P}_{X_t^n})\,\mathrm{d}t \\ \ X_0^{n+1} = x_0. \end{cases}$$

With certain conditions, we can prove that this sequence converges in $L_2(\Omega, \mathcal{C}([0,T], \mathbb{R}^d))$ to the unique solution of (4.2). It is shown in A.5 that the space $L_2(\Omega, \mathcal{C}([0,T], \mathbb{R}^d))$ is a Banach space.

Theorem 5.1 ([BMM19, Theorem 4.1])**.** *Assume that the coefficients b and σ satisfy conditions (L1) and (L2). Then the sequence $((X_t^n)_{t\in[0,T]})_{n=0}^{\infty}$ converges in $L_2(\Omega, \mathcal{C}([0,T], \mathbb{R}^d))$ to a process $X = (X_t)_{t\in[0,T]}$, which is the unique solution to the MVSDE* (4.2).

Before we can prove the theorem above, we prove the following lemma, which is an application of Hölder's inequality 2.9.

Lemma 5.2. *Assume an integrable and measurable function* $f : [0, T] \to \mathbb{R}^d$, $f = (f_1, ..., f_d)$. *Then*

$$\left\| \int_0^t f(s) \, \mathrm{d}s \right\|^2 \leq t \int_0^t \|f(s)\|^2 \, \mathrm{d}s.$$

Proof. By definition we have that

$$\left\| \int_0^t f(s) \, \mathrm{d}s \right\|^2 = \sum_{i=1}^d \left| \int_0^t f_i(s) \, \mathrm{d}s \right|^2 \leq \sum_{i=1}^d \left| \int_0^t |f_i(s) \cdot 1| \, \mathrm{d}s \right|^2.$$

Proposition 2.9 with exponents $p = q = 2$ implies

$$\begin{aligned} \left| \int_0^t |f_i(s) \cdot 1| \, \mathrm{d}s \right|^2 &\leq \left[\left(\int_0^t |f_i(s)|^2 \, \mathrm{d}s \right)^{\frac{1}{2}} \left(\int_0^t |1|^2 \, \mathrm{d}s \right)^{\frac{1}{2}} \right]^2 \\ &= t \int_0^t |f_i(s)|^2 \, \mathrm{d}s \end{aligned}$$

for all $i = 1, ..., d$. Hence

$$\begin{aligned} \left\| \int_0^t f(s) \, \mathrm{d}s \right\|^2 &\leq \sum_{i=1}^d t \int_0^t |f_i(s)|^2 \, \mathrm{d}s = t \int_0^t \left(\sum_{i=1}^d |f_i(s)|^2 \right) \mathrm{d}s \\ &= t \int_0^t \|f(s)\|^2 \, \mathrm{d}s. \end{aligned}$$

$\square$

Now we may give a proof for Theorem 5.1.

Proof of Theorem 5.1. Fix $n \geq 1$. We use the triangle inequality to obtain

$$\begin{aligned} \left\| X_t^{n+1} - X_t^n \right\| =& \Big\| \int_0^t \left[\sigma(s, X_s^n, \mathbb{P}_{X_s^n}) - \sigma(s, X_s^{n-1}, \mathbb{P}_{X_s^{n-1}}) \right] \mathrm{d}B_s \\ &+ \int_0^t \left[b(s, X_s^n, \mathbb{P}_{X_s^n}) - b(s, X_s^{n-1}, \mathbb{P}_{X_s^{n-1}}) \right] \mathrm{d}s \Big\| \\ \leq& \left\| \int_0^t \left[\sigma(s, X_s^n, \mathbb{P}_{X_s^n}) - \sigma(s, X_s^{n-1}, \mathbb{P}_{X_s^{n-1}}) \right] \mathrm{d}B_s \right\| \\ &+ \left\| \int_0^t \left[b(s, X_s^n, \mathbb{P}_{X_s^n}) - b(s, X_s^{n-1}, \mathbb{P}_{X_s^{n-1}}) \right] \mathrm{d}s \right\|. \end{aligned}$$

Then by the Cauchy-Schwarz inequality we have that

$$\begin{aligned}\left\|X_t^{n+1}-X_t^n\right\|^2 \leq & 2\left\|\int_0^t \left[\sigma(s,X_s^n,\mathbb{P}_{X_s^n})-\sigma(s,X_s^{n-1},\mathbb{P}_{X_s^{n-1}})\right]\mathrm{d}B_s\right\|^2 \\ &+2\left\|\int_0^t \left[b(s,X_s^n,\mathbb{P}_{X_s^n})-b(s,X_s^{n-1},\mathbb{P}_{X_s^{n-1}})\right]\mathrm{d}s\right\|^2.\end{aligned}$$

By Lemma 5.2 we have that

$$\begin{aligned}&\left\|\int_0^t \left[b(s,X_s^n,\mathbb{P}_{X_s^n})-b(s,X_s^{n-1},\mathbb{P}_{X_s^{n-1}})\right]\mathrm{d}s\right\|^2 \\ &\leq t\int_0^t \left\|b(s,X_s^n,\mathbb{P}_{X_s^n})-b(s,X_s^{n-1},\mathbb{P}_{X_s^{n-1}})\right\|^2\mathrm{d}s.\end{aligned}$$

Proposition 2.23 gives us the following estimate for the stochastic integral part

$$\begin{aligned}&\mathbb{E}\left(\sup_{t\in[0,T]}\left\|\int_0^t \left[\sigma(s,X_s^n,\mathbb{P}_{X_s^n})-\sigma(s,X_s^{n-1},\mathbb{P}_{X_s^{n-1}})\right]\mathrm{d}B_s\right\|^2\right) \\ &\leq C\mathbb{E}\left(\sqrt{\int_0^T \left\|\sigma(s,X_s^n,\mathbb{P}_{X_s^n})-\sigma(s,X_s^{n-1},\mathbb{P}_{X_s^{n-1}})\right\|^2\mathrm{d}s}\right)^2 \\ &= C\mathbb{E}\int_0^T \left\|\sigma(s,X_s^n,\mathbb{P}_{X_s^n})-\sigma(s,X_s^{n-1},\mathbb{P}_{X_s^{n-1}})\right\|^2\mathrm{d}s\end{aligned}$$

for some absolute constant $C>0$.

Now we have the following inequality

$$\begin{aligned}&\mathbb{E}\left[\sup_{t\in[0,T]}\left\|X_t^{n+1}-X_t^n\right\|^2\right] \\ &\leq 2T\mathbb{E}\int_0^T \left\|b(s,X_s^n,\mathbb{P}_{X_s^n})-b(s,X_s^{n-1},\mathbb{P}_{X_s^n})\right\|^2\mathrm{d}s \\ &\quad+2C\mathbb{E}\int_0^T \left\|\sigma(s,X_s^n,\mathbb{P}_{X_s^n})-\sigma(s,X_s^{n-1},\mathbb{P}_{X_s^{n-1}})\right\|^2\mathrm{d}s.\end{aligned}$$

Using the Lipschitz property of the coefficients b and σ and Lemma 3.4 we continue to

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in[0,T]}\left\|X_t^{n+1}-X_t^n\right\|^2\right]\\
&\leq 2(T+C)L^2\mathbb{E}\int_0^T\left(\left\|X_s^n-X_s^{n-1}\right\|^2+W_2(\mathbb{P}_{X_s^n},\mathbb{P}_{X_s^{n-1}})^2\right)\mathrm{d}s\\
&\leq 4L^2(T+C)\int_0^T\mathbb{E}\left\|X_s^n-X_s^{n-1}\right\|^2\mathrm{d}s\\
&\leq 4L^2(T+C)\int_0^T\mathbb{E}\sup_{r\in[0,s]}\left\|X_r^n-X_r^{n-1}\right\|^2\mathrm{d}s\\
&\leq M_1\int_0^T\mathbb{E}\sup_{r\in[0,s]}\left\|X_r^n-X_r^{n-1}\right\|^2\mathrm{d}s,
\end{aligned}$$

where $M_1>0$ is a constant depending only on the values of C,T and L.

Using similar arguments as earlier and the linear growth condition (L1), we get that

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in[0,T]}\left\|X_t^1-X_t^0\right\|^2\right]&=\mathbb{E}\left[\sup_{t\in[0,T]}\left\|X_t^1-x_0\right\|^2\right]\\
&\leq 2T\int_0^T\left\|b(s,x_0,\mathbb{P}_{X_s^0})\right\|^2\mathrm{d}s+2C\int_0^T\left\|\sigma(s,x_0,\mathbb{P}_{X_s^0})\right\|^2\mathrm{d}s\\
&\leq 2(T+C)\int_0^T K\left(1+\|x_0\|\right)^2\mathrm{d}s\\
&=2(T+C)K\left((1+\|x_0\|)^2\right)T\\
&\leq M_2T,
\end{aligned}$$

where $M_2>0$ is chosen so that the inequality above holds. The choice of M_2 depends on K,C,x and T.

Next we let $f_n(s):=\mathbb{E}\left[\sup_{r\in[0,s]}\|X_r^{n+1}-X_r^n\|^2\right]$ for $n\geq 1$. Now

$$\begin{aligned}
f_n(T)&\leq M_1\int_0^T f_{n-1}(s_{n-1})\,\mathrm{d}s_{n-1}\leq M_1^2\int_0^T\int_0^{s_{n-1}}f_{n-2}(s_{n-2})\,\mathrm{d}s_{n-2}\,\mathrm{d}s_{n-1}\\
&\leq ...\\
&\leq M_1^n\int_0^T\int_0^{s_{n-1}}\cdots\int_0^{s_1}f_0(s_0)\,\mathrm{d}s_0\cdots\mathrm{d}_{s_{n-2}}\,\mathrm{d}_{s_{n-1}}\\
&\leq M_1^n\int_0^T\int_0^{s_{n-1}}\cdots\int_0^{s_1}M_2T\,\mathrm{d}s_0\cdots\mathrm{d}_{s_{n-2}}\,\mathrm{d}_{s_{n-1}}\\
&=\frac{M_1^nM_2T^{n+1}}{n!}\leq\frac{C^{n+1}}{n!},
\end{aligned}$$

where $C := \max\{M_1, M_2\} T$.

Let $m > n$ and $k = m - n$. By triangle inequality we obtain the following estimate

$$\begin{aligned}
\|X^m - X^n\|_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))} &= \left\|X^{n+k} - X^n\right\|_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))} \\
&= \left\|X^{n+k} - X^{n+k-1} - (X^n - X^{n+k-1})\right\|_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))} \\
&\leq \left\|X^{n+k} - X^{n+k-1}\right\|_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))} \\
&\quad + \left\|X^{n+k-1} - X^n\right\|_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))} \\
&\leq \frac{C^{n+k}}{(n+k-1)!} + \left\|X^{n+k-1} - X^n\right\|_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))}.
\end{aligned}$$

By induction on n we obtain

$$\|X^m - X^n\|_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))} \leq \sum_{i=1}^{m-n} \frac{C^{n+i}}{(n+i-1)!}.$$

Letting n and m tend to ∞, we conclude that

$$\|X^m - X^n\|_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))} \to 0,$$

that is, for any $\epsilon > 0$ we can find $N \in \mathbb{N}$ such that

$$\|X^m - X^n\|_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))} < \epsilon$$

for all $m, n \geq N$. Therefore $(X^n)_{n=1}^{\infty}$ is a Cauchy sequence in $L_2(\Omega, \mathcal{C}([0,T], \mathbb{R}^d))$, which is a complete normed space by corollary A.5. It follows that there exists a unique limit $X \in L_2(\Omega, \mathcal{C}([0,T], \mathbb{R}^d))$ such that

$$\|X^n - X\|_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))} \to 0$$

as $n \to \infty$.

Next we need to show that X is a solution to the MVSDE (4.2). We use the same estimates as earlier in this proof to see that

$$\begin{aligned}
&\mathbb{E}\sup_{t\in[0,T]}\left\|\int_0^t \left(b(s,X_s^n,\mathbb{P}_{X_s^n})-b(s,X_s,\mathbb{P}_{X_s})\right)\mathrm{d}s\right\|^2\\
&\quad\leq \mathbb{E}\sup_{t\in[0,T]}\left(t\int_0^t \left\|\left(b(s,X_s^n,\mathbb{P}_{X_s^n})-b(s,X_s,\mathbb{P}_{X_s})\right)\right\|^2\mathrm{d}s\right)\\
&\quad\leq \mathbb{E}\left[T\int_0^T \left\|\left(b(s,X_s^n,\mathbb{P}_{X_s^n})-b(s,X_s,\mathbb{P}_{X_s})\right)\right\|^2\mathrm{d}s\right]\\
&\quad\leq \mathbb{E}\left[2TL^2\int_0^T \left(\|X_s^n-X_s\|^2+W_2(\mathbb{P}_{X_s^n},\mathbb{P}_{X_s})\right)^2\mathrm{d}s\right]\\
&\quad\leq 4TL^2\int_0^T \mathbb{E}\left\|X_s^n-X_s\right\|^2\mathrm{d}s\\
&\quad\leq 4TL^2\int_0^T \mathbb{E}\left[\sup_{r\in[0,T]}\|X_r^n-X_r\|^2\right]\mathrm{d}s\\
&\quad= 4T^2L^2\left\|X^n-X\right\|^2_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))}\to 0
\end{aligned}$$

as n tends to ∞. In a similar way we obtain

$$\begin{aligned}
&\mathbb{E}\sup_{t\in[0,T]}\left\|\int_0^t \left(\sigma(s,X_s^n,\mathbb{P}_{X_s^n})-\sigma(s,X_s,\mathbb{P}_{X_s})\right)\mathrm{d}B_s\right\|^2\\
&\leq C\mathbb{E}\left[\int_0^T \left\|\sigma(s,X_s^n,\mathbb{P}_{X_s^n})-\sigma(s,X_s,\mathbb{P}_{X_s})\right\|^2\mathrm{d}s\right]\\
&\leq 4CL^2\int_0^T \mathbb{E}\left\|X_s^n-X_s\right\|^2\mathrm{d}s\\
&\leq 4CL^2\int_0^T \mathbb{E}\sup_{r\in[0,T]}\|X_r^n-X_r\|^2\mathrm{d}s\\
&\leq 4CL^2T\left\|X^n-X\right\|^2_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))}\to 0
\end{aligned}$$

as n tends to ∞. Combining these two estimates gives us that for $t\in[0,T]$ one has, by definition of X^{n+1},

$$\begin{aligned}
&\left(\mathbb{E}\left\|X_t^{n+1}-x_0-\int_0^t b(s,X_s,\mathbb{P}_{X_s})\,\mathrm{d}s-\int_0^t \sigma(s,X_s,\mathbb{P}_{X_s})\,\mathrm{d}B_s\right\|^2\right)^{\frac{1}{2}}\\
&\leq \left\|\int_0^{\cdot} \left(b(s,X_s^n,\mathbb{P}_{X_s^n})-b(s,X_s,\mathbb{P}_{X_s})\right)\mathrm{d}s\right\|_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))}\\
&\quad+\left\|\int_0^{\cdot} \left(\sigma(s,X_s^n,\mathbb{P}_{X_s^n})-\sigma(s,X_s,\mathbb{P}_{X_s})\right)\mathrm{d}B_s\right\|_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))}\to 0,
\end{aligned}$$

as n tends to ∞. It implies that

$$X_t = x_0 + \int_0^t b(s, X_s, \mathbb{P}_{X_s})\,\mathrm{d}s + \int_0^t \sigma(s, X_s, \mathbb{P}_{X_s})\,\mathrm{d}B_s \tag{5.2}$$

almost surely for any $t \in [0, T]$, and then one gets (5.2) for $t \in [0, T]$ almost surely. Therefore the limit process X is a solution to (4.2). Since assumptions (L1) and (L2) hold for the coefficients σ and b, uniqueness follows from Theorem 4.6. □

Next we consider a straightforward example of how one can use Picard successive approximation to find a solution to the given MVSDE.

Example 5.3. Consider the following MVSDE

$$\begin{cases} \mathrm{d}X_t = \lambda\,\mathrm{d}B_t + \min\left\{\mathrm{e}^T, |\mathbb{E}X_t + 1|\right\}\mathrm{d}t \\ X_0 = 0, \end{cases} \tag{5.3}$$

where $\lambda \in \mathbb{R}$ is a given constant. The coefficient functions clearly satisfy conditions (L1) and (L2). We use Theorem 5.1 to construct a sequence of processes that converges to a unique solution of (5.3).

Our first iteration is

$$X_t^1 = 0 + \int_0^t \lambda\,\mathrm{d}B_s + \int_0^t \min\left\{\mathrm{e}^T, |\mathbb{E}(0) + 1|\right\}\mathrm{d}s = \lambda B_t + t.$$

We continue by computing the next two iterations,

$$X_t^2 = \lambda B_t + \frac{1}{2}t^2 + t$$

and

$$X_t^3 = \lambda B_t + \frac{1}{6}t^3 + \frac{1}{2}t^2 + t.$$

By induction we notice that for arbitrary $n \in \mathbb{N}$ one has

$$X_t^n = \lambda B_t + \sum_{k=1}^{n} \frac{t^k}{k!}.$$

Letting n tend to ∞ we see that

$$X_t^n \to \lambda B_t + \sum_{k=1}^{\infty} \frac{t^k}{k!} = \lambda B_t + \mathrm{e}^t - 1 := X_t,$$

which clearly solves (5.3).

This method cannot always be used to approximate a solution for an MVSDE. We return to example 4.4. We recall that it has no solution, so the iterative method for finding a solution should fail.

Example 5.4. We have the following MVSDE

$$\begin{cases} \mathrm{d}X_t = \mathbb{1}_{\mathbb{Q}}(\mathbb{E}X_t^2)\,\mathrm{d}B_t \\ X_0 = 0. \end{cases}$$

We try to apply Theorem 5.1 to construct a solution.

The first iteration is

$$X_t^1 = \int_0^t \mathbb{1}_{\mathbb{Q}}(\mathbb{E}0^2)\,\mathrm{d}B_s = B_t.$$

Therefore the second one is

$$X_t^2 = \int_0^t \mathbb{1}_{\mathbb{Q}}(\mathbb{E}B_s^2)\,\mathrm{d}B_s = \int_0^t \mathbb{1}_{\mathbb{Q}}(s)\,\mathrm{d}B_s = 0.$$

Next we notice that the third iteration is the same as the first one, so we have that

$$X_t^n = \begin{cases} B_t, & \text{if } n \text{ is odd} \\ 0, & \text{if } n \text{ is even} \end{cases}$$

for $n \in \mathbb{N}$. This sequence does not converge in $L_2(\Omega, \mathcal{C}([0,T], \mathbb{R}))$.

5.2 Stability with respect to the initial condition

In our first stability result we consider solutions of (4.2) for different initial values. If we define a map that maps each initial value to a solution, this map is continuous, assuming that the linear growth and Lipschitz continuity conditions (L1) and (L2) are satisfied.

We denote by $X^x = (X_t^x)_{t\in[0,T]}$ the unique solution to the MVSDE

$$\begin{cases} \mathrm{d}X_t = \sigma(t, X_t, \mathbb{P}_{X_t})\,\mathrm{d}B_t + b(t, X_t, \mathbb{P}_{X_t})\,\mathrm{d}t \\ X_0 = x \end{cases}$$

for the initial value $x \in \mathbb{R}^d$, provided that such a solution exists. We define a map $\varphi : \mathbb{R}^d \to L_2(\Omega, \mathcal{C}([0,T], \mathbb{R}^d)$,

$$\varphi(x) := (X_t^x)_{t\in[0,T]}\,.$$

The next theorem states that, under certain conditions, φ is a continuous map.

Theorem 5.5 ([BMM19, Theorem 5.1]). *Assume that the coefficient functions b and σ satisfy (L1) and (L2). Then the map φ is continuous.*

Proof. First we notice that since the assumptions (L1) and (L2) hold, for all $x \in \mathbb{R}^d$ the unique solution X^x exists.

It is sufficient to show that for any sequence $(x_n)_{n=1}^{\infty}$ in $\mathbb{R}^d$ that converges to the limit $x \in \mathbb{R}^d$ the sequence $(\varphi(x_n))_{n=1}^{\infty}$ converges to $\varphi(x)$ in $L_2(\Omega, \mathcal{C}([0,T], \mathbb{R}^d)$, that is,

$$\|\varphi(x_n) - \varphi(x)\|_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))} = \|X^{x_n} - X^x\|_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))} \to 0$$

as n tends to ∞.

Let $(x_n)_{n=1}^{\infty}$ be a sequence converging to $x \in \mathbb{R}^d$. For a shorter notation, we let $X^n = X^{x_n} = \varphi(x_n)$ and $X = X^x = \varphi(x)$. First we estimate

$$\begin{aligned}
\|X_t^n - X_t\|^2 &= \left\| x_n - x + \int_0^t \big(\sigma(s, X_s^n, \mathbb{P}_{X_s^n}) - \sigma(s, X_s, \mathbb{P}_{X_s})\big)\, \mathrm{d}B_s \right. \\
&\quad \left. + \int_0^t \big(b(s, X_s^n, \mathbb{P}_{X_s^n}) - b(s, X_s, \mathbb{P}_{X_s})\big)\, \mathrm{d}s \right\|^2 \\
&\leq 3\,\|x_n - x\|^2 + 3 \left\| \int_0^t \big(\sigma(s, X_s^n, \mathbb{P}_{X_s^n}) - \sigma(s, X_s, \mathbb{P}_{X_s})\big)\, \mathrm{d}B_s \right\|^2 \\
&\quad + 3 \left\| \int_0^t \big(b(s, X_s^n, \mathbb{P}_{X_s^n}) - b(s, X_s, \mathbb{P}_{X_s})\big)\, \mathrm{d}s \right\|^2,
\end{aligned}$$

where the inequality follows from the Cauchy-Schwarz inequality.

As we did in the proof of Theorem 5.1, we use Lemma 5.2, the Lipschitz condition (L2) and Lemma 3.4 to get

$$\begin{aligned}
&\left\| \int_0^t \big(b(s, X_s^n, \mathbb{P}_{X_s^n}) - b(s, X_s, \mathbb{P}_{X_s})\big)\, \mathrm{d}s \right\|^2 \\
&\leq t \int_0^t \left\| b(s, X_s^n, \mathbb{P}_{X_s^n}) - b(s, X_s, \mathbb{P}_{X_s}) \right\|^2 \mathrm{d}s \\
&\leq t \int_0^t L^2 \left(\|X_s^n - X_s\| + W_2(\mathbb{P}_{X_s^n}, \mathbb{P}_{X_s}) \right)^2 \mathrm{d}s \\
&\leq 2L^2 t \int_0^t \left(\|X_s^n - X_s\|^2 + W_2(\mathbb{P}_{X_s^n}, \mathbb{P}_{X_s})^2 \right) \mathrm{d}s \\
&\leq 2L^2 t \int_0^t \left(\|X_s^n - X_s\|^2 + \mathbb{E}\,\|X_s^n - X_s\|^2 \right) \mathrm{d}s.
\end{aligned}$$

With Proposition 2.23 we obtain

$$\mathbb{E}\left[\sup_{t\in[0,T]}\left\|\int_0^t \left(\sigma(s,X_s^n,\mathbb{P}_{X_s^n})-\sigma(s,X_s,\mathbb{P}_{X_s})\right)\mathrm{d}B_s\right\|^2\right]$$
$$\leq C\mathbb{E}\int_0^T \left\|\sigma(s,X_s^n,\mathbb{P}_{X_s^n})-\sigma(s,X_s,\mathbb{P}_{X_s})\right\|^2\mathrm{d}s$$

for some absolute constant $C>0$. Using the same estimates we used for the coefficient b gives us

$$C\mathbb{E}\int_0^T \left\|\sigma(s,X_s^n,\mathbb{P}_{X_s^n})-\sigma(s,X_s,\mathbb{P}_{X_s})\right\|^2\mathrm{d}B_s$$
$$\leq 2L^2C\mathbb{E}\int_0^T \left(\|X_s^n-X_s\|^2+\mathbb{E}\,\|X_s^n-X_s\|^2\right)\mathrm{d}s$$
$$=4L^2C\int_0^T \mathbb{E}\,\|X_s^n-X_s\|^2\,\mathrm{d}s.$$

Combining these two estimates yields

$$\mathbb{E}\left[\sup_{t\in[0,T]}\|X_t^n-X_t\|^2\right]$$
$$\leq 3\,\|x_n-x\|^2+3T\int_0^T \left\|b(s,X_s^n,\mathbb{P}_{X_s^n})-b(s,X_s,\mathbb{P}_{X_s})\right\|^2\mathrm{d}s$$
$$+3C\mathbb{E}\int_0^T \left\|\sigma(s,X_s^n,\mathbb{P}_{X_s^n})-\sigma(s,X_s,\mathbb{P}_{X_s})\right\|^2\mathrm{d}B_s$$
$$\leq 3\,\|x_n-x\|^2+12(T+C)L^2\int_0^T \mathbb{E}\,\|X_s^n-X_s\|^2\,\mathrm{d}s$$
$$=3\,\|x_n-x\|^2+12(T+C)L^2\int_0^T \mathbb{E}\left[\sup_{r\in[0,s]}\|X_r^n-X_r\|^2\right]\mathrm{d}s.$$

By letting $f_n(t):=\mathbb{E}\left[\sup_{s\in[0,t]}\|X_s^n-X_s\|^2\right]$ for $t\in[0,T]$, we may use Gronwall's inequality with constants $A:=3\,\|x_n-x\|^2$ an $B:=12(T+C)L^2$, to obtain

$$f_n(T)=\mathbb{E}\left[\sup_{s\in[0,t]}\|X_s^n-X_s\|^2\right]=\|X^n-X\|^2_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))}$$
$$\leq A\exp(BT)=3\,\|x_n-x\|^2\exp(12(T+C)L^2T)$$
$$\to 0$$

as n tends to ∞, which completes our proof. □

The preceding theorem has the following application.

Example 5.6. Assume that the Borel-measurable function $\varphi : [0,T]\times\mathbb{R} \to \mathbb{R}$ is bounded and Lipschitz continuous with respect to the second variable. We want to show that the map

$$x \mapsto \left(x + \int_0^{\cdot} \varphi(s,x)\,\mathrm{d}B_s\right)_{t\in[0,T]}$$

is continuous.

We let

$$\sigma(t,x,\mu) := \varphi(s, \int_0^t u\,\mathrm{d}\mu(u))$$

and $b \equiv 0$. The assumptions (L1) and (L2) are clearly satisfied. Now we consider the corresponding MVSDE

$$\begin{cases} \mathrm{d}X_t^x = \varphi(t, \mathbb{E}X_t)\,\mathrm{d}B_t \\ X_0^x = x. \end{cases} \tag{5.4}$$

By Theorem 5.5 the map $x \mapsto (X_t^x)_{t\in[0,T]}$ is continuous. We recall that the expectation of stochastic integral is 0, therefore

$$\mathbb{E}X_t = \mathbb{E}x + \mathbb{E}\int_0^t \varphi(s, \mathbb{E}X_s)\,\mathrm{d}B_s = x,$$

so the solution to 5.4 is

$$X_t^x = x + \int_0^t \varphi(s,x)\,\mathrm{d}B_s.$$

It follows that the map

$$x \mapsto X^x = \left(x + \int_0^{\cdot} \varphi(s,x)\,\mathrm{d}B_s\right)_{t\in[0,T]}$$

is continuous.

5.3 Stability with respect to the coefficients

In some occasions the coefficient functions can be too complicated to efficiently compute or numerically simulate a solution. If we assume that we can find sequences of functions that converge to the coefficients, we could try to approximate the solution by solving the MVSDE with respect to these functions. We want to understand if under some specific assumptions the sequence of these approximated solutions converges to the unique solution of (4.2).

For all $n \in \mathbb{N}$ assume Borel measurable functions

$$b_n : [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$$

and

$$\sigma_n : [0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^{d\times d}.$$

Consider the following MVSDE

$$\begin{cases} \mathrm{d}X_t^n = \sigma_n(t, X_t^n, \mathbb{P}_{X_t^n})\,\mathrm{d}B_t + b_n(t, X_t^n, \mathbb{P}_{X_t^n})\,\mathrm{d}t \\ X_0^n = x_0. \end{cases} \tag{5.5}$$

We want to know whether under certain conditions a solution to (5.5) converges to the unique solution to (4.2) as n tends to ∞. We assume that

$$\int_0^T \left[\sup_{\substack{(x,\mu)\in \\ \mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)}} \left(\|(b_n - b)(s,x,\mu)\|^2 + \|(\sigma_n - \sigma)(s,x,\mu)\|^2 \right) \right] \mathrm{d}s \to 0 \tag{5.6}$$

as n tends to ∞.

Now we may formulate our next stability result. The idea of the proof follows [BMM19, Theorem 6.1], but we have different assumptions.

Theorem 5.7. *Assume that the coefficient functions b_n, b, σ_n and σ satisfy the conditions (L1) and (L2) with uniform constants $K, L > 0$. If* (5.6) *holds, then the sequence $(X^n)_{n=1}^\infty$, where X^n is the unique solution to* (5.5)*, converges to the unique solution of* (4.2) *in $L_2(\Omega, \mathcal{C}([0,T],\mathbb{R}^d))$.*

Proof. First we notice that the assumptions (L1) and (L2) imply that there exists a unique solution to (5.5) and (4.2), which we denote by X^n and X respectively. By the definition of a solution we have

$$\begin{aligned} \|X_t^n - X_t\|^2 &= \left\| \int_0^t \left[b_n(s, X_s^n, \mathbb{P}_{X_s^n}) - b(s, X_s, \mathbb{P}_{X_s}) \right] \mathrm{d}s \right. \\ &\qquad \left. + \int_0^t \left[\sigma_n(s, X_s^n, \mathbb{P}_{X_s^n}) - \sigma(s, X_s, \mathbb{P}_{X_s}) \right] \mathrm{d}B_s \right\|^2 \\ &\le 2 \left\| \int_0^t \left[b_n(s, X_s^n, \mathbb{P}_{X_s^n}) - b(s, X_s, \mathbb{P}_{X_s}) \right] \mathrm{d}s \right\|^2 \\ &\qquad + 2 \left\| \int_0^t \left[\sigma_n(s, X_s^n, \mathbb{P}_{X_s^n}) - \sigma(s, X_s, \mathbb{P}_{X_s}) \right] \mathrm{d}B_s \right\|^2. \end{aligned}$$

We apply Lemma 5.2 and Proposition 2.23 to obtain

$$\mathbb{E}\left[\sup_{t\in[0,T]}\|X_t^n - X_t\|^2\right] \leq 2T\mathbb{E}\int_0^T \left\|b_n(s, X_s^n, \mathbb{P}_{X_s^n}) - b(s, X_s, \mathbb{P}_{X_s})\right\|^2 \mathrm{d}s$$
$$+ 2C\mathbb{E}\int_0^T \left\|\sigma_n(s, X_s^n, \mathbb{P}_{X_s^n}) - \sigma(s, X_s, \mathbb{P}_{X_s})\right\|^2 \mathrm{d}s,$$

where $C > 0$ is an absolute constant. The triangle inequality yields

$$\begin{aligned}
&\left\|b_n(s, X_s^n, \mathbb{P}_{X_s^n}) - b(s, X_s, \mathbb{P}_{X_s})\right\|^2 \\
&\leq \left\|\left[b_n(s, X_s^n, \mathbb{P}_{X_s^n}) - b_n(s, X_s, \mathbb{P}_{X_s})\right] - \left[b(s, X_s, \mathbb{P}_{X_s}) - b_n(s, X_s, \mathbb{P}_{X_s})\right]\right\|^2 \\
&\leq 2\left(\left\|b_n(s, X_s^n, \mathbb{P}_{X_s^n}) - b_n(s, X_s, \mathbb{P}_{X_s})\right\|^2 + \|b_n(s, X_s, \mathbb{P}_{X_s}) - b(s, X_s, \mathbb{P}_{X_s})\|^2\right).
\end{aligned}$$

The same estimate also holds for the functions σ_n and σ. Now

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in[0,T]}\|X_t^n - X_t\|^2\right] \leq{}& 4T\mathbb{E}\int_0^T \left\|b_n(s, X_s^n, \mathbb{P}_{X_s^n}) - b_n(s, X_s, \mathbb{P}_{X_s})\right\|^2 \mathrm{d}s \\
&+ 4C\mathbb{E}\int_0^T \left\|\sigma_n(s, X_s^n, \mathbb{P}_{X_s^n}) - \sigma_n(s, X_s, \mathbb{P}_{X_s})\right\|^2 \mathrm{d}s \\
&+ R_n(T),
\end{aligned}$$

where

$$R_n(t) := 4(t+C)\mathbb{E}\left[\int_0^t \left(\|(b_n - b)(s, X_s, \mathbb{P}_{X_s})\|^2 + \|(\sigma_n - \sigma)(s, X_s, \mathbb{P}_{X_s})\|^2\right) \mathrm{d}s\right]$$

for all $n \in \mathbb{N}$ and $t \in [0, T]$. Since $t \mapsto R_n(t)$ is increasing for all $n \in \mathbb{N}$, we can use assumption (5.6) to conclude that

$$\begin{aligned}
R_n(t) \leq 4(T + C)\mathbb{E}\Bigg[\int_0^T \sup_{(x,\mu)\in\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)} \Bigg(& \|(b_n - b)(s, x, \mu)\|^2 \\
&+ \|(\sigma_n - \sigma)(s, x, \mu)\|^2\Bigg) \mathrm{d}s\Bigg] \to 0
\end{aligned}$$

as n tends to ∞.

The Lipschitz property (L2) and Lemma 3.4 imply that

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{t\in[0,T]} \|X_t^n - X_t\|^2\right] \\
&\leq 4(T+C)\mathbb{E}\int_0^T \left[L(\|X_s^n - X_s\| + W_2(\mathbb{P}_{X_s^n}, \mathbb{P}_{X_s}))\right]^2 \mathrm{d}s + R_n(T) \\
&\leq 8(T+C)L^2 \int_0^T \left(\mathbb{E}\|X_s^n - X_s\|^2 + W_2(\mathbb{P}_{X_s^n}, \mathbb{P}_{X_s})^2\right) \mathrm{d}s + R_n(T) \\
&\leq 16(T+C)L^2 \int_0^T \mathbb{E}\|X_s^n - X_s\|^2 \,\mathrm{d}s + R_n(T) \\
&\leq 16(T+C)L^2 \int_0^T \mathbb{E}\sup_{r\in[0,s]} \|X_r^n - X_r\|^2 \,\mathrm{d}s + R_n(T).
\end{aligned}
$$

We apply Lemma 4.10 to the function

$$
t \mapsto \mathbb{E}\left[\sup_{s\in[0,t]} \|X_s^n - X_s\|^2\right],
$$

where $t \in [0,T]$. By choosing constants $A := R_n(T)$ and $B := 16(T+C)L^2$ we obtain

$$
\|X^n - X\|^2_{L_2(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))} \leq R_n(T)\exp(16(T+C)L^2T) \to 0
$$

as n tends to ∞. This completes the proof. □

In the next example we demonstrate how the preceding theorem can be used to approximate the solutions of an MVSDE.

Example 5.8. Assume bounded and Lipschitz continuous maps

$$
\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}
$$

and

$$
b : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d,
$$

where b is Lipschitz continuous with respect to both variables. Assume that $\|\sigma(x)\| \to 0$ as $\|x\| \to \infty$ and $\|b(x,\mu)\| \to 0$ as $\|x\| \to \infty$ for all $\mu \in \mathcal{P}_2(\mathbb{R})$.

We use notation

$$
B(0,r) := \left\{x \in \mathbb{R}^d \mid \|x\| < r\right\}
$$

and

$$
S(0,r) := \left\{x \in \mathbb{R}^d \mid \|x\| = r\right\}.
$$

We define a map $\pi_r : \mathbb{R}^d \setminus \{0\} \to S(0,r)$ by

$$\pi_r(x) := \frac{rx}{\|x\|},$$

where $r > 0$. Now for all $n \in \mathbb{N}$ we define

$$\sigma_n(x) := \mathbb{1}_{B(0,n)}(x)\sigma(x) + \mathbb{1}_{\mathbb{R}^d \setminus B(0,n)}\sigma(\pi_n(x))$$

and

$$b_n(x,\mu) := \mathbb{1}_{B(0,n)}(x)b(x,\mu) + \mathbb{1}_{\mathbb{R}^d \setminus B(0,n)}b(\pi_n(x),\mu).$$

Since σ and b are bounded, σ_n and b_n are bounded, too, for all $n \in \mathbb{N}$. To check the Lipschitz continuity, let $x \in \mathbb{R}^d \setminus B(0,n)$ and $y \in B(0,n)$. Other cases are trivial. Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$. We denote by $L > 0$ the Lipschitz constant of b. Then

$$\begin{aligned}
\|b_n(x,\mu) - b_n(y,\nu)\| = \|b(\pi_n(x),\mu) - b(y,\nu)\| &= \left\| b\left(\frac{nx}{\|x\|}, \mu\right) - b(y,\nu) \right\| \\
&\leq L\left[\left\|\frac{nx}{\|x\|} - y\right\| + W_2(\mu,\nu)\right] \\
&\leq L\left[\|x-y\| + W_2(\mu,\nu)\right].
\end{aligned}$$

With the same arguments we can prove the Lipschitz continuity of σ_n.

We denote by X^n the unique solution to

$$\begin{cases} \mathrm{d}X_t^n = \sigma_n(X_t^n)\,\mathrm{d}B_t + b_n(X_t^n, \mathbb{P}_{X_t^n})\,\mathrm{d}t \\ X_0^n = x. \end{cases}$$

Next we want to show that (5.6) holds. We see that

$$\sup_{x\in\mathbb{R}^d} \|\sigma_n(x) - \sigma(x)\|^2 = \sup_{x\in\mathbb{R}^d} \left[\mathbb{1}_{\mathbb{R}^d\setminus B(0,n)}(x)\, \|\sigma(x) - \sigma(\pi_n(x))\|^2\right] \to 0$$

as n tends to ∞. In a similar way

$$\sup_{(x,\mu)\in\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)} \|b_n(x,\mu) - b(x,\mu)\|^2 \to 0$$

as n tends to ∞. Hence

$$\begin{aligned}
&\mathbb{E}\left[\int_0^T \sup_{\substack{(x,\mu)\in \\ \mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)}} \left(\|(b_n - b)(x,\mu)\|^2 + \|(\sigma_n - \sigma)(x)\|^2\right) \mathrm{d}s\right] \\
&= T\mathbb{E} \sup_{\substack{(x,\mu)\in \\ \mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)}} \left[\|(b_n - b)(x,\mu)\|^2 + \|(\sigma_n - \sigma)(x)\|^2\right] \to 0
\end{aligned}$$

as $n \to \infty$. Theorem 5.7 implies that $(X^n)_{n=1}^{\infty}$ converges to the unique solution of

$$\begin{cases} \mathrm{d}X_t = \sigma(X_t)\,\mathrm{d}B_t + b(X_t, \mathbb{P}_{X_t})\,\mathrm{d}t \\ X_0 = x. \end{cases}$$

5.4 Stability with respect to the driving process

This far we have considered MVSDEs driven by a Brownian motion. However, when modeling certain phenomena, Brownian motion may not always give the desired behaviour. In this section we consider MVSDEs driven by a continuous martingale $M = (M_t)_{t\in[0,T]} \in \mathcal{M}^{c,0}(\mathbb{R}^d)$, that is

$$\begin{cases} X_t = \sigma(t, X_t, \mathbb{P}_{X_t})\,\mathrm{d}M_t + b(t, X_t, \mathbb{P}_{X_t})\,\mathrm{d}t \\ X_0 = x_0. \end{cases} \tag{5.7}$$

An adapted and continuous process $X = (X_t)_{t\in[0,T]}$ is a solution to (5.7) if it satisfies (Sol1) and (Sol2), and if

$$X_t = x_0 + \int_0^t \sigma(s, X_s, \mathbb{P}_{X_s})\,\mathrm{d}M_s + \int_0^t b(s, X_s, \mathbb{P}_{X_s})\,\mathrm{d}s$$

for $t \in [0, T]$ almost surely, assuming that the integrals exist, see Definition 5.14. Here

$$\int_0^t \sigma(s, X_s, \mathbb{P}_{X_s})\,\mathrm{d}M_s$$

denotes stochastic integral with respect to the continuous martingale M. In one dimension the definition is given in [Mé82, Theorem 18.2]. To multiple dimensions the integral is generalized as in Definition 2.22.

First we notice that we can generalize Theorem 4.6 for this class of MVSDEs.

Theorem 5.9. *Under assumptions (L1) and (L2), the MVSDE* (5.7) *driven by M has a unique solution. Moreover, if $X = (X_t)_{t\in[0,T]}$ is the solution, it holds that*

$$\mathbb{E}\left[\sup_{t\in[0,T]} \|X_t\|^2\right] < \infty.$$

Idea of the proof. We fix $\mu = (\mu_t)_{t\in[0,t]}$ and consider the ordinary SDE

$$\begin{cases} \mathrm{d}X_t = \sigma(t, X_t, \mu_t)\,\mathrm{d}M_t + b(t, X_t, \mu_t)\,\mathrm{d}t \\ X_0 = x_0. \end{cases}$$

The existence and uniqueness of a solution for this SDE is proven in [GK80, chapter 1]. It should be noted that the conditions given for the coefficients are weaker than what we have, but it can be seen that (L1) and (L2) imply these weaker conditions. Now we may adapt the proof of [CD18a, Theorem 4.2.1] by replacing $\mathrm{d}B_t$ with $\mathrm{d}M_t$ and using corresponding theorems for estimates.

Before we may introduce our final stability result, we need a version of Proposition 2.23 for stochastic integrals with respect to a continuous martingale. For this purpose we need to define the *quadratic variation process* of a martingale.

Proposition 5.10 ([Gei19, Proposition 4.4.1]). *Assume a continuous martingale $M = (M_t)_{t\in[0,T]} \in \mathcal{M}^{c,0}(\mathbb{R})$. Then there exists a continuous and adapted process $\langle M \rangle = (\langle M \rangle_t)_{t\in[0,T]}$ such that*

(1) $\langle M \rangle$ is an increasing process starting from 0, that is

$$0 = \langle M \rangle_0 \leq \langle M \rangle_s \leq \langle M \rangle_t$$

for all $s \leq t$.

(2) For every $t \in [0,T]$ and for every partition $0 \leq t_0^n \leq t_1^n \leq ... \leq t_n^n = t$ one has that the sequence

$$\left(\sum_{k=1}^{n} \left(M_{t_n^k} - M_{t_n^{k-1}} \right)^2 \right)_{n=1}^{\infty}$$

converges to $\langle M \rangle_t$ in probability as n tends to ∞.

The process is unique up to indistinguishability.

Definition 5.11 ([Gei19, Proposition 4.4.2]). The process $\langle M \rangle$ in Proposition 5.10 is called quadratic variation of the martingale $M \in \mathcal{M}^{c,0}(\mathbb{R})$.

The quadratic variation has the following property.

Proposition 5.12 ([Mao07, p. 12]). *Let $M \in \mathcal{M}^{c,0}(\mathbb{R})$. Then*

$$M^2 - \langle M \rangle \in \mathcal{M}^{c,0}(\mathbb{R}).$$

In particular $\mathbb{E} \langle M \rangle_t = \mathbb{E} M_t^2$ for all $t \in [0,T]$.

For multidimensional martingales we define the quadratic variation in the following way.

Definition 5.13. Let $M = (M^1, ..., M^d) \in \mathcal{M}^{c,0}(\mathbb{R}^d)$. We define

$$\langle M \rangle := \sum_{i=1}^{d} \langle M^i \rangle .$$

Next we define the class of $\mathcal{L}_2$-processes given a continuous martingale $M \in \mathcal{M}^{c,0}$.

Definition 5.14. Assume $M \in \mathcal{M}^{c,0}$. We denote by $\mathcal{L}_2(M, \mathbb{R})$ the $\mathbb{R}$-valued processes $X = (X_t)_{t\in[0,T]}$ satisfying

$$\mathbb{E} \int_0^t |X_s|^2 \,\mathrm{d} \langle M \rangle_s < \infty. \tag{5.8}$$

for all $t \in [0, T]$. If X is a matrix valued process, then we say that $X \in \mathcal{L}_2(M, \mathbb{R}^{d\times m})$ if $[X^{ij}] \in \mathcal{L}_2(M, \mathbb{R})$ for all $i = 1, ..., d$, $j = 1, ..., m$.

Remark 5.15. If $B = (B_t)_{t\in[0,T]}$ is one-dimensional Brownian motion, then $\langle B \rangle_t = t$ for $t \in [0, T]$ almost surely, see [Gei19, Example 4.4.4]. This coincides with the definition of $\mathcal{L}_2(\mathbb{R})$, so

$$\mathcal{L}_2(\mathbb{R}^{d\times m}) = \mathcal{L}_2(\tilde{B}, \mathbb{R}^{d\times m}),$$

where $\tilde{B}$ is an m-dimensional Brownian motion.

The integral $\int_0^t X_s^2 \,\mathrm{d} \langle M \rangle_s$ is called *Lebesgue-Stieltjes integral* [GG18, Section 5.5.3] with respect to the trajectory $t \mapsto \langle M \rangle_t (\omega)$ for fixed $\omega \in \Omega$, which is continuous and increasing by Proposition 5.10. There exists a unique measure μ in $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that

$$\mu((a, b]) := (\langle M \rangle_b - \langle M \rangle_a)(\omega)$$

for $a < b$. Let $f : [a, b] \to \mathbb{R}$ be an integrable function. We write

$$\int_a^b f(s) \,\mathrm{d} \langle M \rangle_s := \int_{\mathbb{R}} \mathbb{1}_{(a,b]} f(s) \,\mathrm{d}\mu(s).$$

Since $\langle M \rangle$ is a random process, the measure μ might be different for every $\omega \in \Omega$.

$\mathcal{L}_2(M, \mathbb{R})$-processes have the following property.

Proposition 5.16 ([Mé82, Proposition 18.13])**.** *Let $M \in \mathcal{M}^{c,0}(\mathbb{R})$ and $L \in \mathcal{L}_2(M, \mathbb{R})$. Then*

$$\left\langle \int_0^{\cdot} L_s \,\mathrm{d}M_s \right\rangle_t = \int_0^t L_s^2 \,\mathrm{d} \langle M \rangle_s .$$

Now we may formulate a more general version of Burkholder-Davis-Gundy inequality.

Theorem 5.17 (Burkholder-Davis-Gundy, [RY99, Chapter IV, Theorem 4.1])**.** *Let $M \in \mathcal{M}^{c,0}(\mathbb{R})$. Then there exists constants c_p and C_p such that for all $t \in [0,T]$ one has*

$$c_p \mathbb{E}\left[\langle M \rangle_t^{\frac{p}{2}}\right] \le \mathbb{E}\left[\sup_{s\in[0,t]} |M_s|^p\right] \le C_p \mathbb{E}\left[\langle M \rangle_t^{\frac{p}{2}}\right].$$

As a consequence we get the following estimate for integrals with respect to a continuous martingale.

Corollary 5.18. *Let $X = [X^{ij}]_{i,j=1,\dots,d} \in \mathcal{L}_2(M, \mathbb{R}^{d\times d})$ and $M \in \mathcal{M}^{c,0}(\mathbb{R}^d)$. Then there exists a constant $C > 0$ such that*

$$\mathbb{E}\left[\sup_{t\in[0,T]} \left\| \int_0^t X_r \,\mathrm{d}M_r \right\|^2\right] \le C\mathbb{E}\left[\int_0^T \|X_s\|^2 \,\mathrm{d}\langle M\rangle_s\right].$$

Proof. If $d = 1$, we use Proposition 5.16 and Proposition 5.17 with $p = 2$ to see that

$$\mathbb{E}\left[\sup_{t\in[0,T]} \left| \int_0^t X_s \,\mathrm{d}M_s \right|^2\right] \le C_2\mathbb{E}\left[\int_0^T X_s^2 \,\mathrm{d}\left\langle M\right\rangle_s\right]. \tag{5.9}$$

For $d > 1$ we recall that, by definition,

$$\left\| \int_0^t X_s \,\mathrm{d}M_s \right\|^2 = \sum_{k=1}^d \left| \sum_{j=1}^d \int_0^t X_s^{kj} \,\mathrm{d}M_s \right|^2.$$

We fix $k = 1, ..., d$ and see that

$$\left| \sum_{j=1}^d \int_0^t X_s^{kj} \,\mathrm{d}M_s \right|^2 \le d \sum_{j=1}^d \left| \int_0^t X_s^{kj} \,\mathrm{d}M_s \right|^2.$$

Then

$$\left\| \int_0^t X_s \,\mathrm{d}M_s \right\|^2 \le d \sum_{k=1}^d \sum_{j=1}^d \left| \int_0^t X_s^{kj} \,\mathrm{d}M_s \right|^2.$$

Next we apply (5.9) to each term of the sum to get that

$$\begin{aligned}
\mathbb{E}\left[\sup_{t\in[0,T]}\left\|\int_0^t X_s \,\mathrm{d}M_s\right\|^2\right] &\leq d\sum_{k=1}^{d}\sum_{j=1}^{d}\mathbb{E}\left[\sup_{t\in[0,T]}\left|\int_0^t X_s^{kj}\,\mathrm{d}M_s^j\right|^2\right] \\
&\leq dC_2\sum_{k=1}^{d}\sum_{j=1}^{d}\mathbb{E}\left[\int_0^T \left|X_s^{kj}\right|^2 \mathrm{d}\left\langle M^j\right\rangle_s\right] \\
&\leq dC_2\sum_{j=1}^{d}\mathbb{E}\left[\int_0^T \left(\sum_{k=1}^{d}\left|X_s^{kj}\right|^2\right) \mathrm{d}\left\langle M^j\right\rangle_s\right] \\
&\leq dC_2\mathbb{E}\left[\int_0^T \left(\sum_{j,k=1}^{d}\left|X_s^{kj}\right|^2\right) \mathrm{d}\left(\sum_{j=1}^{d}\left\langle M^j\right\rangle_s\right)\right] \\
&\leq dC_2\mathbb{E}\left[\int_0^T \|X_s\|^2 \,\mathrm{d}\left\langle M\right\rangle_s\right].
\end{aligned}$$

□

Another lemma wee need is the *stochastic Gronwall inequality.*

Lemma 5.19 (Stochastic Gronwall inequality, [Mé82, Lemma 29.1])**.** *Assume a continuous martingale $M = (M_t)_{t\in[0,T]} \in \mathcal{M}^{c,0}(\mathbb{R}^d)$ satisfying*

$$\langle M\rangle_T = \sup_{t\in[0,T]} \langle M\rangle_t \leq C$$

for some constant $C > 0$ almost surely. Let $\varphi = (\varphi_t)_{t\in[0,T]}$ be a real-valued, adapted and increasing process. Assume that there exists positive constants K and ρ such that for all $t \in [0,T]$ one has

$$\mathbb{E}\varphi_t \leq K + \rho\mathbb{E}\int_0^t \varphi_s \,\mathrm{d}\langle M\rangle_s .$$

Then

$$\mathbb{E}\varphi_T \leq KD,$$

where

$$D := 2\sum_{k=1}^{\lfloor 2\rho C\rfloor} (2\rho C)^k \in (0,\infty).$$

Now we may formulate our next stability result. Let $M \in \mathcal{M}^{c,0}(\mathbb{R}^d)$. Denote by $X = (X_t)_{t\in[0,T]}$ the unique solution to the MVSDE driven by M.

Assume a sequence of continuous martingales $(M^n)_{n=1}^{\infty}$ with $M_0^n \equiv 0$ for all $n \in \mathbb{N}$. We consider the following MVSDE

$$\begin{cases} X_t^n = \sigma(t, X_t^n, \mathbb{P}_{X_t^n}) \,\mathrm{d}M_t^n + b(t, X_t^n, \mathbb{P}_{X_t^n}) \,\mathrm{d}t \\ X_0^n = x_0, \end{cases} \tag{5.10}$$

We assume that the following assumptions are satisfied:

(D1) There exists a constant $\lambda > 0$ such that

$$\langle M^n \rangle_T + \langle M \rangle_T \leq \lambda$$

almost surely for all $n \in \mathbb{N}$.

(D2) It holds that

$$\mathbb{E} \langle M^n - M \rangle_T \to 0$$

as n tends to ∞.

Under these assumptions, we have the following result. The theorem and its proof are inspired by [BMM19, Theorem 7.1], but we have different assumptions.

Theorem 5.20. *Assume that the coefficient functions b and σ are bounded and satisfy (L1) and (L2). If the assumptions (D1) and (D2) are satisfied, then $(X^n)_{n=1}^{\infty}$, where X^n is the unique solution to* (5.10)*, converges to the unique solution of* (5.7) *in $L_2(\Omega, \mathcal{C}([0,T], \mathbb{R}^d))$.*

Proof. We denote by X the unique solution of (5.7). Since b and σ satisfy (L1) and (L2), the uniqueness and existence of X^n and X holds. We notice that since the coefficients are bounded, the integrals exist and thus we may proceed with our proof.

By the definition of a solution we have

$$\begin{aligned} \|X_t^n - X_t\|^2 &= \bigg\| \int_0^t \sigma(s, X_s^n, \mathbb{P}_{X_s^n}) \,\mathrm{d}M_s^n - \int_0^t \sigma(s, X_s, \mathbb{P}_{X_s}) \,\mathrm{d}M_s \\ &\quad + \int_0^t \big[b(s, X_s^n, \mathbb{P}_{X_s^n}) - b(s, X_s, \mathbb{P}_{X_s}) \big] \,\mathrm{d}s \bigg\|^2 \\ &\leq 2 \bigg\| \int_0^t \sigma(s, X_s^n, \mathbb{P}_{X_s^n}) \,\mathrm{d}M_s^n - \int_0^t \sigma(s, X_s, \mathbb{P}_{X_s}) \,\mathrm{d}M_s \bigg\|^2 \\ &\quad + 2 \bigg\| \int_0^t \big[b(s, X_s^n, \mathbb{P}_{X_s^n}) - b(s, X_s, \mathbb{P}_{X_s}) \big] \,\mathrm{d}s \bigg\|^2 . \end{aligned}$$

By Lemma 5.2 and the Lipschitz property (L2) we obtain

$$\begin{aligned}
&\left\| \int_0^t \left[b(s, X_s^n, \mathbb{P}_{X_s^n}) - b(s, X_s, \mathbb{P}_{X_s}) \right] \mathrm{d}s \right\|^2 \\
&\leq t \int_0^t \left\| b(s, X_s^n, \mathbb{P}_{X_s^n}) - b(s, X_s, \mathbb{P}_{X_s}) \right\|^2 \mathrm{d}s \\
&\leq 2tL^2 \int_0^t \left(\|X_s^n - X_s\|^2 + W_2(\mathbb{P}_{X_s^n}, \mathbb{P}_{X_s})^2 \right) \mathrm{d}s.
\end{aligned}$$

Then by Lemma 3.4 we have that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{t\in[0,T]} \left\| \int_0^t \left[b(s, X_s^n, \mathbb{P}_{X_s^n}) - b(s, X_s, \mathbb{P}_{X_s}) \right] \mathrm{d}s \right\|^2 \right] \\
&\leq 4TL^2 \int_0^T \mathbb{E} \sup_{r\in[0,s]} \|X_r^n - X_r\|^2 \,\mathrm{d}r.
\end{aligned}$$

For a shorter notation we let $\varphi_n(t) := \sigma(t, X_t^n, \mathbb{P}_{X_t^n})$ for all $n \in \mathbb{N}$ and $\varphi(t) := \sigma(t, X_t, \mathbb{P}_{X_t})$. We estimate

$$\begin{aligned}
&\left\| \int_0^t \varphi_n(s) \,\mathrm{d}M_s^n - \int_0^t \varphi(s) \,\mathrm{d}M_s \right\|^2 \\
&= \left\| \int_0^t (\varphi_n(s) - \varphi(s)) \,\mathrm{d}M_s^n + \int_0^t \varphi(s) \,\mathrm{d}(M_s^n - M_s) \right\|^2 \\
&\leq 2 \left(\left\| \int_0^t (\varphi_n(s) - \varphi(s)) \,\mathrm{d}M_s^n \right\|^2 + \left\| \int_0^t \varphi(s) \,\mathrm{d}(M_s^n - M_s) \right\|^2 \right).
\end{aligned}$$

We fix $u \in [0, T]$ and apply Corollary 5.18 on both terms to obtain

$$\begin{aligned}
&2\mathbb{E}\left[\sup_{t\in[0,u]} \left(\left\| \int_0^t (\varphi_n(s) - \varphi(s)) \,\mathrm{d}M_s^n \right\|^2 + \left\| \int_0^t \varphi(s) \,\mathrm{d}(M_s^n - M_s) \right\|^2 \right) \right] \\
&\leq 2C\mathbb{E}\left[\int_0^u \|\varphi_n(s) - \varphi(s)\|^2 \,\mathrm{d}\langle M^n \rangle_s + \int_0^u \|\varphi(s)\|^2 \,\mathrm{d}\langle M^n - M \rangle_s \right]
\end{aligned}$$

for some constant $C > 0$.

Next we apply Lipschitz property (L2) and Lemma 3.4 to get

$$
\begin{aligned}
&\mathbb{E}\int_0^u \|\varphi_n(s)-\varphi(s)\|^2 \,\mathrm{d}\langle M^n\rangle_s \\
&= \mathbb{E}\int_0^u \left\|\sigma(s,X_s^n,\mathbb{P}_{X_s^n})-\sigma(s,X_s,\mathbb{P}_{X_s})\right\|^2 \mathrm{d}\langle M^n\rangle_s \\
&\le 2L^2\int_0^u \mathbb{E}\left(\|X_s^n-X_s\|^2+W_2(\mathbb{P}_{X_s^n},\mathbb{P}_{X_s})^2\right)\mathrm{d}\langle M^n\rangle_s \\
&\le 4L^2\int_0^u \mathbb{E}\,\|X_s^n-X_s\|^2\,\mathrm{d}\langle M^n\rangle_s \\
&\le 4L^2\int_0^u \mathbb{E}\left[\sup_{r\in[0,s]}\|X_r^n-X_r\|^2\right]\mathrm{d}\langle M^n\rangle_s .
\end{aligned}
$$

By assumptions σ is bounded, hence there exists a constant $B>0$ such that

$$
\begin{aligned}
\mathbb{E}\int_0^u \|\varphi(s)\|^2\,\mathrm{d}\langle M^n-M\rangle_s &= \mathbb{E}\int_0^u \|\sigma(s,X_s,\mathbb{P}_{X_s})\|^2\,\mathrm{d}\langle M^n-M\rangle_s \\
&\le \mathbb{E}\int_0^u B^2\,\mathrm{d}\langle M^n-M\rangle_s \\
&= \mathbb{E}B^2\left(\langle M^n-M\rangle_u-\langle M^n-M\rangle_0\right) \\
&= B^2\mathbb{E}\langle M^n-M\rangle_u \\
&\le B^2\mathbb{E}\langle M^n-M\rangle_T .
\end{aligned}
$$

We combine our estimates to see that for all $u\in[0,T]$ it holds that

$$
\begin{aligned}
\mathbb{E}\left[\sup_{t\in[0,u]}\|X_t^n-X_t\|^2\right] \le\ & 8CL^2\mathbb{E}\left[\int_0^u \sup_{r\in[0,s]}\|X_r^n-X_r\|^2\,\mathrm{d}\langle M^n\rangle_s\right] \\
&+2CB^2\mathbb{E}\langle M^n-M\rangle_T \\
&+8TL^2\int_0^T \mathbb{E}\sup_{r\in[0,s]}\|X_r^n-X_r\|^2\,\mathrm{d}s.
\end{aligned}
$$

We apply Lemma 5.19 to the process $\left(\sup_{s\in[0,t]}\|X_s^n-X_s\|^2\right)_{t\in[0,T]}$, which implies that

$$
\mathbb{E}\left[\sup_{t\in[0,T]}\|X_t^n-X_t\|^2\right] \le D\mathbb{E}\langle M^n-M\rangle_T + D\int_0^T \mathbb{E}\sup_{r\in[0,s]}\|X_r^n-X_r\|^2\,\mathrm{d}s
$$

where $D>0$ is a constant.

To complete the proof we apply Lemma 4.10 and assumption (D2) to see that

$$\mathbb{E}\left[\sup_{t\in[0,T]} \|X_t^n - X_t\|^2\right] \leq D\mathbb{E}\left\langle M^n - M\right\rangle_T \mathrm{e}^{DT} \to 0$$

as n tends to ∞. □

We demonstrate Theorem 5.20 in the context of ordinary stochastic differential equations.

Example 5.21. Assume a sequence of uniformly bounded $\mathcal{L}_2(\mathbb{R})$-processes $(L^n)_{n=1}^{\infty}$ and a process $L \in \mathcal{L}_2(\mathbb{R})$. Assume that

$$|L_t^n - L_t| \underset{a.s.}{\to} 0$$

for all $t \in [0,T]$. Define $M_t^n := \int_0^t L_s^n \,\mathrm{d}B_s$ and $M_t := \int_0^t L_s \,\mathrm{d}B_s$. We have that

$$\langle M\rangle_T = \int_0^T |L_s|^2 \,\mathrm{d}s \leq TB^2 \quad \text{and} \quad \langle M^n\rangle_T = \int_0^T |L_s^n|^2 \,\mathrm{d}s \leq TB^2$$

for all $n \in \mathbb{N}$, where $B > 0$ is a constant.

Using our assumptions we see that

$$\langle M^n - M\rangle_T = \int_0^T |L_s^n - L_s|^2 \,\mathrm{d}s \to 0$$

almost surely as n tends to ∞. It follows that

$$\mathbb{E}\left\langle M^n - M\right\rangle_T \to 0$$

as n tends to ∞.

Now we may use Theorem 5.20 to deduce that the sequence of unique solutions to the following ordinary stochastic differential equations

$$\begin{cases} X_t^n = X_t^n \,\mathrm{d}M_t^n \\ X_0^n = 1 \end{cases}$$

converges in $L_2(\Omega, \mathcal{C}([0,T],\mathbb{R})$ to the unique solution to

$$\begin{cases} X_t = X_t \,\mathrm{d}M_t \\ X_0 = 1. \end{cases}$$

We recall that we may write

$$\int_0^t X_s^n \,\mathrm{d}M_s^n = \int_0^t X_s^n L_s^n \,\mathrm{d}B_s.$$

We apply Thoerem 2.24 with the function $f(t,x) := \log(x)$ to the process X_t^n to obtain that

$$\begin{aligned}\log(X_t^n) &= \log(1) + \int_0^t L_s^n \,\mathrm{d}B_s - \frac{1}{2}\int_0^t L_s^2 \,\mathrm{d}s \\ &= M_t^n - \frac{1}{2}\langle M\rangle_t^n .\end{aligned}$$

Hence

$$X_t^n = \exp\left(M_t^n - \frac{1}{2}\langle M\rangle_t^n\right).$$

By Theorem 5.20 this implies that

$$\begin{aligned}\mathbb{E}\Bigg[\sup_{t\in[0,T]}\Bigg|&\exp\left(\int_0^t L_s^n \,\mathrm{d}B_s - \frac{1}{2}\int_0^t (L_s^n)^2 \,\mathrm{d}s\right) \\ &- \exp\left(\int_0^t L_s \,\mathrm{d}B_s - \frac{1}{2}\int_0^t (L_s)^2 \,\mathrm{d}s\right)\Bigg|^2\Bigg] \to 0\end{aligned}$$

as n tends to ∞.

Appendix A $L_p(\Omega, E)$ spaces

Assume a real vector space V and a map $\|\cdot\| : V \to [0, \infty)$. The map $\|\cdot\|$ is a norm, if the following properties are satisfied:

(1) $\|v\| = 0$ if and only if $v = 0$.

(2) $\|\cdot\|$ satisfies the *triangle inequality*, that is, for all $u, v \in V$ one has that

$$\|u + v\| \leq \|u\| + \|v\| .$$

(3) $\|\cdot\|$ is *positive homogeneous*, that is, for all $\lambda \in \mathbb{R}$ and $v \in V$ one has

$$\|\lambda v\| = |\lambda| \, \|v\| .$$

The pair $(V, \|\cdot\|)$ is called *normed space*. If the normed space is complete, that is, every Cauchy sequence converges to a limit in V, then it is called *Banach space*. The space $(V, \|\cdot\|)$ is *separable*, if there exists a countable dense subset $B \subseteq V$.

Definition A.1 ([Egg84, page 3])**.** Let $p \in [1, \infty)$. Assume a measure space $(\Omega, \mathcal{F}, \mu)$ and a normed space $(E, \|\cdot\|_E)$. Denote by $\mathcal{L}_p(\Omega, E)$ the set of all $(\mathcal{F}, \mathcal{B}(E))$-measurable maps $\varphi : \Omega \to E$ satisfying

(1) $\|\varphi(\omega)\|_E < \infty$ for all $\omega \in \Omega$,

(2) $\|\varphi\|_{L_p(\Omega,E)} := \left(\int_\Omega \|\varphi(\omega)\|_E^p \, \mathrm{d}\mu\right)^{\frac{1}{p}} < \infty$.

It should be noted that $\|\cdot\|_{L_p(\Omega,E)}$ is only a semi-norm in $\mathcal{L}_p(\Omega, E)$ since for any map $\varphi : \Omega \to E$ with $\mathbb{P}(\varphi = 0) = 1$, that is, $\varphi(\omega) = 0$ almost everywhere, we have $\|\varphi\|_{L_p(\Omega,E)} = 0$, but not necessarily $\varphi \equiv 0$. However, by using certain equivalence classes we may construct a normed space.

Definition A.2. Define an equivalence relation $\sim$ in $\mathcal{L}_p(\Omega, E)$ by letting $\varphi \sim \psi$ if and only if $\mathbb{P}(\varphi = \psi) = 1$. Let

$$L_p(\Omega, E) := \{[\varphi] \mid \varphi \in \mathcal{L}_p(\Omega, E)\} ,$$

where $[\varphi]$ denotes the equivalence class consisting of all the maps $\psi \in \mathcal{L}_p(\Omega, E)$ with $\varphi \sim \psi$.

In this section our goal is to show that if $(\Omega, \mathcal{F}, \mu)$ is a probability space, then the space $L_p(\Omega, E)$ with the norm $\|\cdot\|_{L_p(\Omega,E)}$ is a Banach space for all $p \in [1, \infty)$. First we define what we mean by convergence with respect to the norm $\|\cdot\|_{L_p(\Omega,E)}$.

Definition A.3. Assume a sequence of $(E, \mathcal{B}(E))$-measurable maps $(f_n)_{n=1}^{\infty}$, $f_n : \Omega \to E$ with $f_n \in L_p(\Omega, E)$. We say that the sequence $(f_n)_{n=1}^{\infty}$ converges in $L_p(\Omega, E)$ to the limit $f \in L_p(\Omega, E)$ if

$$\|f_n - f\|_{L_p(\Omega,E)} \to 0$$

as n tends to ∞.

The following theorem is mentioned in [Egg84, Theorem I.1.2.3] without proof. A proof for a special case, where $E = \mathbb{R}$ is given in [GG18, Proposition 6.3.4]. The proof we give here adapts that proof.

Theorem A.4. *Let $p \in [1, \infty)$. Let $(E, \|\cdot\|_E)$ be a separable Banach space. Then the space $(L_p(\Omega, E), \|\cdot\|_{L_p(\Omega,E)})$ is a Banach space.*

Proof. We prove the theorem in two steps. First we prove that $(L_p(\Omega, E), \|\cdot\|_{L_p(\Omega,E)})$ is a normed space. In the second part we prove the completeness, meaning that any Cauchy sequence converges to a limit in $(L_p(\Omega, E), \|\cdot\|_{L_p(\Omega,E)})$.

Part 1: Norm properties

We show that the norm $\|\cdot\|_{L_p(\Omega,E)}$ has the following three properties:

(1) Assume that $\|\varphi\| = 0$. Then $\|\varphi(\omega)\|_E = 0$ almost everywhere. Since $\|\cdot\|_E$ is a norm, we get that $\varphi(\omega) = 0$ almost everywhere, that is, $\varphi \in [0]$, which is the zero vector in $L_p(\Omega, E)$.

  Next assume that $\varphi = 0$. In this context it means that $\varphi(\omega) = 0$ almost everywhere, and therefore $\|\varphi\|_{L_p(\Omega,E)} = 0$.

(2) **Triangle inequality.** Let $\varphi, \psi \in L_p(\Omega, E)$. Then

$$\begin{aligned}
\|\varphi + \psi\|_{L_p(\Omega,E)} &= \left(\mathbb{E}\,\|\varphi + \psi\|_E^p\right)^{\frac{1}{p}} \\
&\le \left(\mathbb{E}\,[\|\varphi\|_E + \|\psi\|_E]^p\right)^{\frac{1}{p}} \\
&\le \left(\mathbb{E}\,\|\varphi\|_E^p\right)^{\frac{1}{p}} + \left(\mathbb{E}\,\|\psi\|_E^p\right)^{\frac{1}{p}} \\
&= \|\varphi\|_{L_p(\Omega,E)} + \|\psi\|_{L_p(\Omega,E)},
\end{aligned}$$

  where the second inequality follows from Proposition 2.10.

(3) **Positive homogeneous.** Let $\lambda \in \mathbb{R}$ and $\varphi \in L_p(\Omega, E)$. Then

$$\begin{aligned}
\|\lambda\varphi\|_{L_p(\Omega,E)} &= \left(\mathbb{E}\,\|\lambda\varphi\|_E^p\right)^{\frac{1}{p}} = \left(\mathbb{E}\left[|\lambda|^2\,\|\varphi\|_E^p\right]\right)^{\frac{1}{p}} \\
&= |\lambda|\left(\mathbb{E}\,\|\varphi\|_E^p\right)^{\frac{1}{p}} = |\lambda|\,\|\varphi\|_{L_p(\Omega,E)}.
\end{aligned}$$

Part 2: Completeness

Let $(\varphi_n)_{n=1}^{\infty}$ be a Cauchy sequence in $L_p(\Omega, E)$. Fix $\epsilon > 0$. Then there exists $N_\epsilon \in \mathbb{N}$ depending on the choice of ϵ such that

$$\|\varphi_m - \varphi_n\|_{L_p(\Omega,E)}^p = \mathbb{E}\,\|\varphi_m - \varphi_n\|_E^p \leq \epsilon \tag{A.1}$$

for all $m, n \geq N_\epsilon$. Note that we take the norm to the p-th power, which does not affect the convergence. We use Chebyshev's inequality [GG18, Proposition 5.10.1] to see that for any $\lambda > 0$ we have

$$\mathbb{P}\left(\left\{\omega \in \Omega \;\middle|\; \|\varphi_m(\omega) - \varphi_n(\omega)\|_E > \lambda\right\}\right) \leq \frac{1}{\lambda^p}\mathbb{E}\,\|\varphi_m - \varphi_n\|_E^p \leq \epsilon.$$

Hence $(\varphi_n)_{n=1}^{\infty}$ is a Cauchy sequence in probability, and this implies in particular that there exists a limit $\varphi : \Omega \to E$ such that $(\varphi_n)_{n=1}^{\infty}$ converges to φ in probability. This lets us choose a subsequence $(\varphi_{n_k})_{k=1}^{\infty}$ such that the subsequence converges almost surely to the limit φ with respect to the norm $\|\cdot\|_E$, that is

$$\mathbb{P}\left(\left\{\omega \in \Omega \mid \|\varphi_{n_k}(\omega) - \varphi(\omega)\|_E \not\to 0 \text{ as } k \to \infty\right\}\right) = 0.$$

We apply Fatou's lemma [GG18, Proposition 5.4.4] to obtain

$$\begin{aligned}\mathbb{E}\,\|\varphi - \varphi_n\|_E^p &= \mathbb{E}\left[\liminf_{k\to\infty} \|\varphi_{n_k} - \varphi_n\|_E^p\right] \\ &\leq \liminf_{k\to\infty} \mathbb{E}\,\|\varphi_{n_k} - \varphi_n\|_E^p \leq \epsilon\end{aligned}$$

for $n \geq N_\epsilon$. Hence $(\varphi_n)_{n=1}^{\infty}$ converges to a limit in $L_p(\Omega, E)$.

□

We recall that the space of continuous functions, $\mathcal{C}([0,T], \mathbb{R}^d)$ with the norm $\|\cdot\|_\infty$ defined by

$$\|f\|_\infty := \sup_{t\in[0,T]} \|f(t)\|$$

is a separable Banach space. Now we obtain the following result.

Corollary A.5. *For all $T > 0$ the space $L_p(\Omega, \mathcal{C}([0,T], \mathbb{R}^d))$ with the norm*

$$\|X\|_{L_p(\Omega,\mathcal{C}([0,T],\mathbb{R}^d))} := \left(\mathbb{E}\,\|X\|_\infty^p\right)^{\frac{1}{p}} = \left(\mathbb{E}\sup_{t\in[0,T]} \|X(t,\cdot)\|^p\right)^{\frac{1}{p}}$$

for $X : [0,T] \times \Omega \to \mathbb{R}^d$, is a Banach space.

Proof. Follows directly from Theorem A.4 when $E = \mathcal{C}([0,T], \mathbb{R}^d)$. □

References

[BMM19] K. Bahlali, M. Mezerdi, and B. Mezerdi. Stability of mckean–vlasov stochastic differential equations and applications. February 2019. arXiv:1902.03478 [math.PR].

[Car16] R. Carmona. *Lectures on BSDEs, Stochastic Control, and Stochastic Differential Game with Financial Applications*. Society of Industrial and Applied Mathematics, 2016.

[CD18a] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications I*. Springer International Publishing AG, 2018.

[CD18b] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications II*. Springer International Publishing AG, 2018.

[Egg84] L. Egghe. *Stopping time techniques for analysts and probabilists.* Cambridge University press, 1984.

[Gei19] S. Geiss. *Stochastic Differential Equations.* University of Jyväskylä, Department of Mathematics and Statistics, 2019.

[GG18] C. Geiss and S. Geiss. *An Introduction to Probability Theory.* University of Jyväskylä, Department of Mathematics and Statistics, 2018.

[GK80] I. Gyöngy and N. V. Krylov. On stochastic equations with respect to semimartingales i. *Stochastics*, 4:1–21, September 1980.

[JP04] R. Jarrow and P. Protter. A short history of stochastic integration and mathematical finance: The early years, 1880–1970. *A Festschrift for Herman Rubin Institute of Mathematical Statistics Lecture Notes – Monograph Series*, 45:75–91, 2004.

[Kac56] M. Kac. Foundations of kinetic theory. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 3: Contributions to Astronomy and Physics*, pages 171–197, Berkeley, Calif., 1956. University of California Press.

[LL07] J. Lasry and P. Lions. Mean field games. *Japanese journal of mathematics*, 2:229–260, February 2007.

[Mé82] M. Métivier. *Semimartingales - a course in stochastic processes.* De Gruyter Studies in Mathematics, 1982.

[Mao07] X. Mao. *Stochastic differential equations and applications.* Woodhead Publishing, 2 edition, 2007.

[McK66] H. P. McKean. A class of markov processes associated with non-linear parabolic equations. *Proceedings of the National Academy of Sciences of the United States of America*, 56(6):1907–1911, 1966.

[Mey09] P. Meyer. Stochastic processes from 1950 to the present. *Journal Electronique d'Histore des Probabilités et de la Statistique*, 5(1), June 2009.

[PCMM15] F. Pegoraro, F. Califano, G. Manfredi, and P. Morrison. Theory and applications of the vlasov equation. *The European Physical Journal D*, 69, 02 2015.

[Rud70] W. Rudin. *Real and Complex Analysis.* McGraw-Hill Inc., 1970.

[RY99] D. Revuz and M. Yor. *Continuous Martingales and Brownian Motion.* Springer, third edition, 1999.

[Sob91] K. Sobczyk. *Stochastic Differential Equations.* Springer, 1991.

[Vil06] C. Villani. *Optimal transport, old and new.* Springer, 2006.