This thesis has been submitted in fulfilment of the requirements for a postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights, which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given.
Estimates of space derivatives for functions of non-autonomous and McKean-Vlasov processes. Application to uniform weak error bounds induced by the approximating subsampled particle system

Maria Lefter

THE UNIVERSITY of EDINBURGH

HERIOT WATT UNIVERSITY

MIGSAA

Doctor of Philosophy
University of Edinburgh
April 17, 2023
Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

(Maria Lefter)
April 17, 2023
Abstract

This thesis is split in three parts, all of which obtain derivative estimates for the solution to the backward Kolmogorov equation associated to diverse stochastic process and study their application to uniform weak error bounds. More specifically, throughout the document we consider the function of time and space obtained after evaluating a fixed test function at a process with deterministic initial condition (space component) and governed by a predetermined SDE (time component). This function satisfies the so-called backward Kolmogorov equation and we refer to it as the backward Kolmogorov function.

The first part of the thesis studies space derivatives estimates for the backward Kolmogorov function associated to non-autonomous SDEs. It presents sufficient conditions for derivatives' decay in the time component and shows how these can be obtained from customary assumptions used in finite time PDE analysis when combined with monotonicity.

The second part applies these results to the scenario where the underlying stochastic process is a McKean-Vlasov process and presents complementary results which lead to uniform in time gradient estimates of the backward Kolmogorov function. The obtained bounds do not rely on results available in the literature and are based on lax regularity assumptions of the coefficients and again monotonicity conditions.

The third part develops a more intuitive, although more restricted, way of obtaining decaying in time derivative estimates of the backward Kolmogorov function associated to a McKean–Vlasov process. Moreover, it uses them to derive uniform weak error estimates for the time discretization of the subsampled particle system approximating a McKean-Vlasov process.
Lay summary

Many fields study some “characteristic” of an interacting (over time) ensemble of $N$ particles/individuals. When their behaviour is irregular and highly volatile or it cannot be measured explicitly, the branch of mathematics concerned with modelling these phenomena is stochastic analysis. This will consist on seeing the particle as an accumulation of possibilities(samples) and not as a certain event. Combining these, $N$ interacting particles whose behaviour cannot be explicitly measures is then expressed mathematically as a system of $N$ stochastic differential equations (SDEs). Since big systems are hard to deal with, mathematical analysis is performed instead on the limiting SDE as the number of particles $N$ goes to infinity. The resulting object is a McKean–Vlasov process and the “characteristic” averaged across the ensemble is represented by a function evaluated at it. This is just one possible motivation for studying McKean-Vlasov processes, which are more broadly defined as SDEs whose coefficients depend on the law/distribution of the solution itself and are relevant on their own.

However, in practice, McKean–Vlasov SDEs are rarely solvable explicitly. Hence, we flip the problem once more and approximating efficiently its solution by an associated particle system – from which we can sample – becomes highly relevant. Part of this thesis analyses the subsampling technique: only a small, random batch of particles interact at each instant and in the next instant they are shuffled again, in order to reduce computational cost when sampling from it. The accuracy of such model, meaning the error induced by this extra randomness, is studied from the weak error’s point of view. This is, from the point of view of distance to the “characteristic” of the original McKean–Vlasov process.

This dissertation is dedicated to the study of the subsampling technique and the necessary and more general technical study of the derivatives of such a “characteristic” of the targeted process when seen as a function of time and space.
I want to convey my gratitude to my supervisor Dr. David Šiška for his patience in sharing his knowledge and expertise while guiding my research. I would like to extend my gratitude to my collaborator Dr. Lukasz Szpruch for his valuable feedback and contributions. Without their relentless support this endeavour would have been impossible. In addition, I am extremely grateful to my second supervisor Dr. Michela Ottobre for the encouragement she offered me over the past three years.

I acknowledge MIGSA for providing funding and extensive training opportunities which broadened my knowledge in diverse fields of mathematics, fostered my research communication skills and showed me that mathematics have a greater impact than I ever imagined. Apart from the broad selection of talks organised by both the University of Edinburgh and Heriot–Watt University, I was given to opportunity to attend events such as the “9th International Colloquium on Backward Stochastic Differential Equations”, the “Connections between interacting particle dynamics and data science” Conference and the “13th European Summer School in Financial Mathematics”. I offer my sincere gratitude to the organisers. I also had a great pleasure being part of the “Machine Learning” reading group organised by my two supervisors and the “Infinite dimensional analysis” reading group organised by Dr. Ottobre, which gave me a thorough understanding of topics undergoing intense study and broadened my knowledge of the field.

I also send my appreciation to my Undergraduate supervisor Dr. Josè–Luis Torrea Hernandez for his words of comfort during the PhD but also for encouraging my scientific enthusiasm through his professionalism and enjoyment towards mathematics.

Finally, I would like to thank enormous my family, and in particular my beloved mother, for their unwavering support through the hardest moments and for their accompanying excitement during the good ones. This list would not be complete without the friends with who I shared the PhD experience and made it so much better: special mention to Alina, Kieran, Tiffany and Fabian.
## Contents

Abstract iv  
Lay summary v  
Acknowledgments vi  

1 Introduction 1  
1.1 Fundamentals of Analysis and Probability 1  
1.1.1 Probability concepts 1  
1.1.2 Itô calculus 3  
1.1.3 Existence and uniqueness of solution to SDEs 6  
1.1.4 Semigroup theory 8  
1.2 Classical Analysis and Probability theorems 9  
1.3 Kolmogorov equations 12  
1.3.1 Derivation 12  
1.3.2 Results from PDEs literature 16  
1.3.3 Ergodicity 18  
1.4 Measure space calculus 20  
1.4.1 McKean–Vlasov SDEs 22  
1.5 Particle systems 25  
1.5.1 Particle system approximation 26  
1.5.2 Propagation of chaos 27  
1.6 Euler scheme 29  
1.7 Notation 30  
1.7.1 Notation 30  
1.8 Overview 31  

2 Decaying derivative estimates for functions of solutions to non-autonomous SDEs 33  
2.1 Introduction 33  
2.2 Main result: Derivative with respect to the initial condition 34  
2.3 Application to non-autonomous SDEs 37  
2.3.1 Finite time estimates on derivatives of transition densities 39  
2.3.2 Convergence to a measure under Lyapunov condition 43  
2.3.3 Result for autonomous SDEs 46  
2.4 Application to non-autonomous SDE decaying to an autonomous one 48
3 Gradient estimates for linear functions of solutions to McKean–Vlasov SDEs 55
  3.1 Introduction ................................................................. 55
  3.1.1 Recall: Gradient in the first space component ......................... 57
  3.2 Derivative with respect to the initial data of the auxiliary backward Kolmogorov function ................................................. 60
  3.2.1 Smoothness and integrability of derivative of the density w.r.t starting point ......................................................... 62
  3.2.2 Convergence to invariant measure ........................................ 70
  3.3 Derivative with respect to the parameter of the auxiliary backward Kolmogorov function .................................................... 75
  3.4 Gradient of the backward Kolmogorov function for a McKean–Vlasov SDE ................................................................. 80

4 Weak error of subsampled approximation of McKean-Vlasov SDEs 82
  4.1 Introduction ................................................................. 82
  4.1.1 PDE theory for weak error analysis ........................................ 86
  4.1.2 Assumptions ................................................................. 87
  4.1.3 Motivational example ..................................................... 91
  4.2 Subsampling discussion .................................................... 93
  4.2.1 Subsampling in continuous time ........................................... 94
  4.2.2 Dependence on the size of the batches .................................. 98
  4.3 Derivatives decay for solutions to the backward Kolmogorov equation ................................................................. 103
  4.3.1 Derivative processes ....................................................... 105
  4.3.2 Space derivatives of the backward Kolmogorov function ............... 114
  4.4 Euler Scheme approximation of the subsampled system .................... 116
  4.5 Further research .............................................................. 125

Appendices 126

A Appendix Chapter 2 127
  A.1 Non–homogeneous Chapman–Kolmogorov identity ....................... 127

B Appendix Chapter 3 128
  B.1 Moments estimates ............................................................ 128
  B.2 Examples with decaying bounds for $\partial_y^2 V$ .......................... 133
  B.2.1 Lyapunov assumptions ...................................................... 133
  B.2.2 Veretennikov–Khasminskii assumptions .................................. 135
  B.2.3 Another set of Lyapunov conditions ...................................... 136

C Appendix Chapter 4 139
  C.1 Strong propagation of chaos for the subsampled system ................. 139
  C.2 Moment bounds ............................................................... 141
  C.3 Glossary of derivative processes .......................................... 147
  C.4 Weak error for the Euler scheme of a McKean–Vlasov SDE .............. 151
  C.5 Weak error induced by the discretization of the full particle system .... 156

Bibliography 162
Chapter 1

Introduction

This first chapter aims to provide the necessary background for the remainder of the thesis. It covers basic concepts from analysis and probability and introduces the reader to the theory of PDEs, SDEs and the relation in between them induced by the Kolmogorov equations. More detail is included on PDEs, ergodicity, calculus on measure spaces and McKean–Vlasov SDEs in order to prepare the reader for the results borrowed from the literature which are used in the subsequent chapters. Every result presented has a reference (in the title) to the book or article which contains its proof.

1.1 Fundamentals of Analysis and Probability

For preparing this section we mainly followed [1] and [2].

We use the notation \( \mathbb{R} \) for the real numbers and \( \mathbb{R}_+ \) for the set of the non-negative ones among those. Moreover, let \( E \) be a vector space equipped with a norm \( |\cdot| \). In our case, we mostly consider the Euclidean space \( \mathbb{R}^d \) (for some \( d \in \mathbb{N} \)) with the associated Euclidean norm \( \mathbb{R}^d \ni x = (x_1,\ldots,x_d) \mapsto |x|_2 = (x_1^2 + \ldots + x_d^2)^{1/2} \). Since this norm is born from the inner product \( \mathbb{R}^d \times \mathbb{R}^d \ni (x,y) \mapsto xy = \langle x,y \rangle = (x_1 - y_1)^2 + \ldots + (x_d - y_d)^2 \) and \( (\mathbb{R}^d,\langle \cdot,\cdot \rangle) \) is a Hilbert space, this choice gives us access to a wide set of analysis tools. Another useful vector space is \( \mathbb{R}^d \) equipped with a \( p \)-norm \( (p \geq 1) \), i.e. for \( x = (x_1,\ldots,x_d) \in \mathbb{R}^d \), the norm is \( |x|_p = (x_1^p + \ldots + x_d^p)^{1/p} \).

1.1.1 Probability concepts

**Definition 1.1.1.** Given a set \( \Omega \), a \( \sigma \)-algebra (or \( \sigma \)-field ) on \( \Omega \), denoted by \( \mathcal{F} \), is a subset of the power set of \( \Omega \) such that:

- \( \Omega \in \mathcal{F} \);
- \( A \in \mathcal{F} \) implies that \( \Omega - A \in \mathcal{F} \);
- \( A_1, A_2, \ldots \in \mathcal{F} \) implies that \( \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{F} \).

We say that the tuple \( (\Omega,\mathcal{F}) \) is a measurable space and any set \( A \in \mathcal{F} \) is said to be measurable.

**Definition 1.1.2.** A measure on a measurable space \( (\Omega,\mathcal{F}) \) is any function \( \mu : \Omega \to [0,\infty] \) such that
• \( \mu(\emptyset) = 0; \)
• For any collection of pairwise disjoint sets \( A_1, A_2, \ldots \in \mathcal{F} \), we have \( \mu( \bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu(A_n) \).

Moreover, if \( \mu(\Omega) = 1 \) then we say that \( \mu \) is a probability measure and the custom notation is then \( \mu = \mathbb{P} \).

The most common \( \sigma \)-algebra on \( E \) is the so called Borel \( \sigma \)-algebra \( (\mathcal{B}(E)) \), which is the smallest \( \sigma \)-algebra such that all open sets are measurable.

**Definition 1.1.3.** An \( E \)-valued random variable (r.v.) on the probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \) is a function \( X : (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{B}(E)) \) that satisfies for any \( A \in \mathcal{B}(E) \) that \( X^{-1}(A) \in \mathcal{F} \).

Given a \( \mathbb{R}^d \)-valued random variable \( X \), it induces naturally another measure onto \( \mathbb{R}^d \): for any \( A \in \mathbb{R}^d \), the law of the random variable \( X \) is defined as \( \mathcal{L}(X)(A) = \mathbb{P}(X \in A), \forall A \in \mathcal{B}(\mathbb{R}^d) \).

In the particular case that \( A_x = (-\infty, x_1] \times \ldots \times (-\infty, x_d] \) for some \( x \in \mathbb{R}^d \) we can define the so called cumulative distribution function \( F_X(x) = \mathcal{L}(X)(A_x) \).

Moreover, if \( \mathcal{L}(X) \) is absolutely continuous with respect to the Lebesgue measure, by the Radón Nikodym Theorem, there exists a probability density function \( p_X : \mathbb{R}^d \rightarrow \mathbb{R} \) such that \( \mathcal{L}(X)(A) = \int_A p_X(dx) \).

Other relevant spaces will be: \( \mathcal{B}(E) \), the space of \( E \)-Borel–measurable (i.e. the \( \sigma \)-algebra on \( E \) is the Borel one) functions, the space of \( p \)-continuously differentiable functions over \( I \) (elements from \([0, 1)\times \ldots \times [0, 1)\)) polynomially growing \( \mathbb{E} \)-denoted by \( C^{p_1, p_2}(I \times \mathcal{B}(E)) \).

A \( b \) as the subindex, i.e. \( C^{p_1, p_2}([0, T) \times \mathcal{B}(E)) \), will mean that the functions and all \( p_1 \)-order derivatives in time (elements from \([0, T)\)) and \( p_2 \)-order derivatives in space (elements from \( E \)) are bounded. A particularly important role is played by \( m \)–order \( (m \in \mathbb{N}) \) polynomially growing functions \( B_m \), i.e. \( \phi \in \mathcal{B}(E) \) such that there exists \( C > 0 \) such that \( |\phi(x)| \leq C(1+|x|^m) \), \( \forall x \in E \).

Moreover, the space of probability measures on \( E \) is represented by \( \mathcal{P}(E) \). Finally, for any strongly measurable function \( \phi \in \mathcal{B}(E) \) and \( q \) a measure on \( E \), we say \( \phi \in L^p(q) \) whenever \( \int_E |\phi(x)|^p d(q(dx)) < \infty \). The measure is only omitted when we refer to the Lebesgue measure.

Other important concepts are that of expectation of the random variable \( X \) given by \( \mathbb{E}[X] = \int_E x \mathcal{L}(X)(dx) \), variance of the same given by \( \text{Var}(X) = \mathbb{E}[(X-\mathbb{E}[X])^2] \) and in general any \( p \geq 1 \)–moment given by \( \mathbb{E}[|X|^p] \).

We can now define the concept of conditional expectation.

**Definition 1.1.4.** Given a probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \), a sub–\( \sigma \)-algebra \( \mathcal{I} \subset \mathcal{F} \) and \( X : \Omega \rightarrow \mathbb{R}^n \) an integrable random variable such that \( \mathbb{E}[|X|] < \infty \), the conditional expectation of \( X \) given \( \mathcal{I} \) is the random variable \( \mathbb{E}[X|\mathcal{I}] \) satisfying:

• \( \mathbb{E}[X|\mathcal{I}] \) is \( \mathcal{I} \)-measurable;
• \( \mathbb{E}[X|\mathcal{I}] \in L^1(\Omega) \);
• For every event \( F \in \mathcal{F} \) and its index function \( 1_F \),

\[ \mathbb{E}[1_F \mathbb{E}[X|\mathcal{I}]] = \mathbb{E}[X1_F]. \]

With respect to distances in any Euclidean space, \( |\cdot| \) is used as an arbitrary norm (since they are all equivalent in finite dimensions and our estimates would only change up to a constant), only the trace norm denoted by \( \text{tr}(\cdot) \) being mentioned independently. We next introduce metrics on measure spaces.
Theorem 6.18]. A Polish space is a separable completely metrizable topological space. Hence, the union of all of them. In other words, \( \sigma \)-field is 

\[
\mathcal{F} = \bigcup_{\omega \in \Omega} \mathcal{B}(\mathbb{R})
\]

where \( \pi \) belongs to the product \( \mathcal{P}(\Omega \times \mathbb{R}) \). Then, the \( \sigma \)-algebra generated from \( \mathcal{F} \) is denoted by \( \mathcal{F} = \sigma(\mathcal{F}) \).

Continuing with the notation listing, for any matrix \( M \in \mathbb{R}^{d \times d} \), we denote its transpose by \( M^* \), its determinant by \( \det(M) \) and its Frobenius norm by \( \| M \| = \sqrt{\text{tr}(MM^*)} \). However, notice that given the finite dimensional setup, all the matrix norms are equivalent and we will denote a generic matrix norm by \( \| \cdot \| \). Next, let us mention that throughout the dissertation \( C > 0 \) is a constant changing value from line to line. Despite not being an appreciated practice and since we do not care about the value of the constants but rather their dependence on the parameters, the author considered simplifying the notation this way. Finally, let us mention that if they exist, we denote the \( n \)-th derivative of the function \( f: \mathbb{R}^d \to \mathbb{R} \) by \( \partial \alpha f(x) = \partial_{x_1}^{\alpha_1} \ldots \partial_{x_d}^{\alpha_d} f(x) \), where \( x = (x_i)_{i=1}^d \in \mathbb{R}^d \) and \( \alpha = (\alpha_i)_{i=1}^d \in \mathbb{N}^d \) is such that \( |\alpha| = n \).

1.1.2 Itô calculus

A stochastic process is a mathematical model for the occurrence, at each moment after the initial time, of a random phenomenon. Namely, we define the following.

Definition 1.1.6. Given a probability space \( (\Omega, \mathcal{F}, \mathbb{P}) \), a measurable space \( (S, \Sigma) \) and an interval \( T \subseteq \mathbb{R} \), a collection of \( S \)-valued random variables \( X = (X_t)_{t \in T} \) is a stochastic process.

Moreover, we say that the stochastic process \( X \) is measurable if for every \( A \in \Sigma \), the set \( \{(t, \omega) : X_t(\omega) \in A \} \) belongs to the product \( \sigma \)-algebra \( B(T) \otimes \mathcal{F} \).

The temporal feature of stochastic processes suggest a flow of time, which is why at each present time one can wonder how much an observer of the process knows compared to a point in the past or in the future. All this information is contained within a filtration, the concept we define next.

Definition 1.1.7. Given a measurable space \( (\Omega, \mathcal{F}) \), a filtration is a non-decreasing family of \( \sigma \)-algebras \( (\mathcal{F}_t)_{t \geq 0} \), sub-\( \sigma \)-algebras of \( \mathcal{F} \), where \( \mathcal{F}_\infty \) is the sigma-algebra generated from the union of all of them. In other words,

\[
\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}, \quad \forall 0 \leq s \leq t; \quad \text{and} \quad \mathcal{F}_\infty = \sigma(\bigcup_{t \geq 0} \mathcal{F}_t).
\]
The next step is to quantify the gain or loss of information over time. And for that, a particularly useful concept is the following.

**Definition 1.1.8.** A random variable $T$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0})$ is called a stopping time if for any $t \geq 0$, the sets $\{T \leq t\}$ are $\mathcal{F}_t$–measurable.

A random process in the same space, $(X_t)_{t \geq 0}$, is called adapted to the filtration if for any $t \geq 0$, $X_t$ are $\mathcal{F}_t$–measurable.

**Definition 1.1.9.** Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, equipped with a filtration $(\mathcal{F}_t)_{t \geq 0}$ and a random process $(X_t)_{t \geq 0}$ such that $X_t: \Omega \rightarrow \mathbb{R}^d$ for any $t \geq 0$. We say that $(X_t)_{t \geq 0}$ is a submartingale (respectively, supermartingale) with respect to $\mathcal{F}$ if for every $t \geq 0$,

- $X_t$ is $\mathcal{F}_t$–measurable,
- $\mathbb{E}[|X_t|] < \infty$,
- and for every $0 \leq s \leq t$, we have that $\mathbb{P}$ – a.s $\mathbb{E}[X_t | \mathcal{F}_s] \geq X_s$ (respectively, $\mathbb{E}[X_t | \mathcal{F}_s] \leq X_s$).

We shall say that $(X_t)_{t \geq 0}$ is a martingale if it is both a submartingale and a supermartingale and a semimartingale if it satisfies the definition of possibly just one of them.

Moreover, we say that a process $(X_t)_{t \geq 0}$ is a local martingale if there exists an increasing sequence of stopping times $(\tau_n)_{n \in \mathbb{N}}$ such that $\tau_n \rightarrow \infty$ as $n \rightarrow \infty$ and the stopped process $(X_{\tau_n})_{n \in \mathbb{N}}$ is a martingale.

Intuitively, a martingale conserves the information over time.

The Brownian Motion, also known as Wiener process, is arguably the most important stochastic process and it is in fact a martingale.

**Definition 1.1.10.** Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the $d$–dimensional Brownian Motion (Weiner process) is a stochastic process $B := (B_t)_{t \geq 0}$, $B_t: \Omega \rightarrow \mathbb{R}^d$, satisfying:

- $B_0 = 0$;
- $B$ has (a.s.) continuous paths (i.e. for every $\omega \in \Omega$ fixed, $B(. \omega)$ are continuous functions of $t \in [0, \infty]$);
- For any sequence $0 \leq t_0 \leq t_1 \leq \ldots \leq t_n$, for integers $n \geq 2$, $W_{t_1} - W_{t_0}, \ldots, W_{t_n} - W_{t_{n-1}}$ are independent random variables;
- For every $0 \leq s \leq t < \infty$, the increments $B_t - B_s$ are distributed as a $d$–dimensional normal $\mathcal{N}(0, (t-s)I_d)$, where $I_d$ is the $d$–dimensional identity matrix.

Born to model the movement of pollen particles in water and credited to the botanist Robert Brown in 1827, it lies at the bottom of the Itô’s calculus. Its chaotic behaviour is its strength for modelling particle movement or stock prices but also its weakness for an easy mathematical analysis.

We first mention informally what we mean by Itô’s integral $\int_0^T X_t(\omega)dB_t(\omega)$ since, due to its unbounded variation on any finite time interval, this integral cannot be understood in the usual Lebesgue-Stieltjes sense. Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a filtration $(\mathcal{F}_t)$, and an $\mathcal{F}_t$–Wiener martingale $B$, one starts by defining Itô’s integral with a
class of integrands called simple processes. Then, by using an approximating argument, one extends the definition for any $\mathcal{F}_t$–adapted processes whose square's integral in time over the interval $(0, \infty)$ has finite expectation. A final step is based on another approximating argument in order to define Itô’s integral for any $\mathcal{F}_t$–adapted processes $(X_t)$ satisfying for all $s > 0$ that $\int_0^s X_t^2 \, dt < \infty$, almost surely. A relevant property of this stochastic integral process $(I_s := \int_0^s X_t \, dB_t)_{t \geq 0}$ is that it is a continuous local martingale with respect to the filtration $(\mathcal{F}_s)$. If moreover $\int_0^T \mathbb{E}[X_t^2] \, dt < \infty$ for all $s \geq 0$, then it is a $(\mathcal{F}_t)$–martingale and in particular it satisfies $\mathbb{E}[I_s] = 0$ and $\mathbb{E}[|I_t|^2] = \mathbb{E}\left[\int_0^T X_t^2 \, dt\right]$. The latter is known by the name of Itô’s isometry.

**Definition 1.1.11.** Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a Brownian motion $(B_t)_{t \geq 0}$. We call a process $(X_t)_{t \in [0,T]}$ a Itô process, if almost surely

$$X_t^\xi = \xi + \int_0^t b(X_s^\xi) \, ds + \int_0^t \sigma(X_s^\xi) \, dB_s, \quad \forall t \in [0, T],$$

where $\xi$ is an $\mathcal{F}_0$–measurable random variable,

- $b: \mathbb{R}^d \to \mathbb{R}^d$ is such that $(b_t(X_t))_{t \in [0,T]}$ is an $\mathcal{F}_t$–adapted process satisfying that $\int_0^T b_t(X_t) \, dt < \infty$ almost surely and for all $s \in (0, T)$ and $i = 1, \ldots, d$;

- and $\sigma: \mathbb{R}^d \to \mathbb{R}^{d \times m}$ with $\sigma = (\sigma^{ij})_{i=1,\ldots,d}^{j=1,\ldots,m}$ is such that $(\sigma^{ij}(X_t))_{t \in [0,T]}$ is an $\mathcal{F}_t$–adapted process satisfying that $\int_0^T (\sigma^{ij}(X_t))^2 \, dt < \infty$ almost surely and for all $s \in (0, T)$ and $i = 1, \ldots, d; j = 1, \ldots, m$.

We then say that $X^\xi = (X_t^\xi)_{t \geq 0}$ satisfies the stochastic differential equation (SDE):

$$dX_t^\xi = b(X_t^\xi) \, dt + \sigma(X_t^\xi) \, dB_t, \quad \forall t \in [0, \infty); \quad X_0^\xi = \xi.$$ 

The reciprocal, i.e. find a solution to a given SDE, admits various interpretations.

**Definition 1.1.12.** Consider the stochastic differential equation (1.1.2).

- Given as well the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and fixed the Brownian motion $(B_t)_{t \geq 0}$ and initial condition $\xi$, a strong solution to SDE (1.1.2) is a process $X = (X_t)_{t \geq 0}$ with continuous sample paths and the following properties:

  - $X$ is adapted to the filtration $(\mathcal{F}_t)$ constructed from $\mathcal{F}$ such that $\mathcal{F}_0$ contains the necessary null sets and for all $t > 0$, $\mathcal{F}_t = \sigma\left(\xi, B_s: s \in [0, t]\right) \cup \mathcal{F}_0$;

  - $\mathbb{P}(X_0 = \xi) = 1$

  - for all $t \geq 0$, and $i = 1, \ldots, d; j = 1, \ldots, m$ we have $\mathbb{P}\left(\left\{\int_0^t |b(X_s^\xi)| + |\sigma^{ij}(X_s^\xi)|^2 \, ds < \infty\right\}\right) = 1$;

  - (1.1.1) holds almost surely.

- A weak solution to (1.1.2) is a triple $(X, B), (\Omega, \mathcal{F}, \mathbb{P})$ and $(\mathcal{F}_t)$ where:
Theorem 1.1.14. (Existence and uniqueness of solution to SDEs [5, Theorem 1.2]) Consider the SDE given by the generalisation of (1.1.2) holds almost surely.

The analogous of the Chain Rule was generalised to d-dimensional problems and even to convex but not necessarily differentiable functions (see [4, Theorem 5.3.1] and [1, Theorem 6.22]).

Theorem 1.1.13 (Itô’s formula). Let X be a d-dimensional Itô process with stochastic differential given by the generalisation of (1.1.2) with time-dependent coefficients:

\[ dX^i_t = b(t, X^i_t)dt + \sigma(t, X^i_t)dB_t, \quad \forall t \in [0, T] \]  

(1.1.3)

and let \( f \in C^{1,2}([0, T] \times \mathbb{R}^d) \). Then, \( (f(t, X_t))_{t \in [0, T]} \) is a Itô process with stochastic differential

\[ df(t, X_t) = \partial_t f(t, X_t) + \sum_{i=1}^d b_i(t, X_t)\partial_{x_i} f(t, X_t)dt + \sum_{i,j=1}^d \frac{1}{2}(\sigma\sigma^*)^{ij}(t, X_t)\partial^2_{x_i x_j} f(t, X_t)dt + \sum_{i=1}^d \sum_{k=1}^m \sigma^{ik}(t, X_t)\partial_{x_i} f(t, X_t)dB^k_t. \]  

(1.1.4)

1.1.3 Existence and uniqueness of solution to SDEs

This section is devoted to the most classical (strong) existence and uniqueness theorems for solution to (1.1.2). Moreover, it provides results regarding the existence and regularity of transition densities.

Theorem 1.1.14. (Existence and uniqueness of solution to SDEs [5, Theorem 1.2]) Consider the SDE given by (1.1.2) such that \( \xi \) is a \( \mathcal{F}_0 \)-measurable r.v., the coefficients are jointly continuous and for any finite time \( T > 0 \) and \( R > 0 \), we have the following local coercivity (local one-sided Lipschitz) and monotonicity (global one-sided linear growth) conditions satisfied: for all \( |x|, |y| \leq R \) and \( t \geq 0 \)

\[ 2\langle x - y, b(t, x) - b(t, y) \rangle + ||\sigma(t, x) - \sigma(t, y)||^2 \leq K_t(R)|x - y|^2, \]
\[ 2\langle x, b(t, x) \rangle + ||\sigma(t, x)||^2 \leq K_t(1+|x|^2), \]

where \( K_t(R) \) are certain adapted, non-negative process satisfying \( \int_0^T K_t(R)dt < \infty, \forall R, T > 0 \). Then, there exist a unique (a.s.) solution to (1.1.2).

In particular, the above conditions are satisfied for globally Lipschitz coefficients such that \( \int_0^T |b(t, 0)| + ||\sigma(t, 0)||^2 dt < \infty \) for all \( T \geq 0 \).

A generalisation of the above is done by using Lyapunov techniques to guarantee uniqueness of solutions even when the local Lipschitz conditions do not hold.

6
Theorem 1.1.15. (Existence and uniqueness of solution to SDEs [6, Theorem 3.5]) Given the SDE (1.1.2) and associated infinitesimal generator $C^2(\mathbb{R}^d) \ni f \rightarrow Lf = b \partial_x f + 1/2 \text{tr}(\sigma \sigma^* f)$, suppose that the coefficients $b, \sigma$ are continuous and for every compact $K \subset [0, \infty) \times \mathbb{R}^d$ we have $C > 0$ such that

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

$$|b(t, x)| + ||\sigma(t, x)|| \leq C(1 + |x|) \quad \forall (t, x), (t, y) \in K.$$

Moreover, suppose that there exists a non–negative function $W \in C^2(\mathbb{R}^d)$ such that for some constant $c > 0$,

$$LW \leq cW \quad \text{and} \quad \inf_{|x| \leq R} W(x) \to \infty \text{ as } R \to \infty.$$

Then:

- For every random variable $\xi$ independent of the Brownian Motion $B$, there exists a solution $X$ to (1.1.2) which is an a.s. continuous stochastic process, unique up to indistinguishability.

- This solution is a Markov process whose transition probability function is defined by $\mathcal{L}(X_t^{s,x})(A) = \mathbb{P}(X_t^{s,x} \in A)$ for every $0 \leq s \leq t$, $x \in \mathbb{R}^d$ and $A \in B(\mathbb{R}^d)$.

- If the coefficients $b, \sigma$ are independent of $t$, then the transition probability function of the corresponding Markov process is time–homogeneous.

- This process also satisfies, for all $t \geq 0$ and $\xi$ chosen as above, that there exists $c > 0$ such that

$$\mathbb{E}[W(t, X_t^{s,x})] \leq \mathbb{E}[W(0, \xi)]e^{-ct}.$$

See [7] for further recent results on existence and uniqueness of SDEs.

Two of the most fundamental properties which stochastic processes can have are the Markovian and Martingale properties (see [5, Theorem 2.10] and [1, Proposition 5.4.2]).

After establishing existence and uniqueness results, the next step is the analysis of properties of the solution. A Markovian process $(X_t)_{t \geq 0}$ admits a transition density if for each $x \in \mathbb{R}^d, 0 \leq s \leq t < \infty$, there exists a measurable function $\mathbb{R}^d \ni y \mapsto p_t^s(x, y)$ such that for any $A \in \mathcal{F}_s$:

$$p_t^s(x, A) := \mathcal{L}(X_t^{s,x})(A) = \int_A p_t^s(x, y)dy = \int_A \mathcal{L}(X_t^{s,x})(dy).$$

The following theorem gives conditions for its existence (when taking $\mu = \mathcal{L}(X_t^0)$).

Theorem 1.1.16. [8, Theorem 1.5.2] Let $O \subset \mathbb{R}^d$ be an open set and consider the process $X$ satisfying the SDE (1.1.2), suppose that the matrix $\sigma \sigma^*(x)$ is non–negative for every $x \in O$ and $\mu$ is a finite Borel measure on $O$ (i.e he defined in all opens sets of $O$ and such that $\mu(O) < \infty$ but possibly signed) such that for any compact $K \subset O$ and $i, j = 1,...,d, \int_K (\sigma \sigma^*)(ij)(x)d\mu(x) < \infty$ (locally integrable) and for some $C > 0$

$$\int_O (\sigma \sigma^*)(ij)\partial^2_{xixj}\phi d\mu \leq C(\sup_O |\phi| + \sup_O |\partial_x \phi|),$$

for all non–negative $\phi \in C^\infty_0(O)$. Then, the following assertions hold:
• If \( \mu \) is non–negative, then \( \det(\sigma \sigma^*)^{1/d} \mu \) has a \( d' \)–locally integrable density in the dual space, where \( d' = \frac{d}{d-1} \).

• If \( \sigma \sigma^* \) is locally Hölder continuous and \( \det(\sigma \sigma^*) > 0 \), then \( \mu \) has a \( r \)–locally integrable density in the dual space for every \( r \in [1, d') \).

However, the most popular result on this line is the Hörmander condition, which not only gives existence of density, but also its regularity. But before that let us define a Lie bracket operation.

**Definition 1.1.17.** Given \( V, W \in C^1(\mathbb{R}^d; \mathbb{R}^d) \), let \([V, W] \) denote their Lie bracket, another vector field defined by
\[
[V, W](x) = \partial_x V(x) W(x) - \partial_x W(x) V(x),
\]
where \( \partial_x V(x) \) denotes the Fréchet derivative of \( V \) at \( x \in \mathbb{R}^d \) and vice–versa.

**Theorem 1.1.18.** (Hörmander’s theorem [9, Theorem 6.1]) Consider the Itô stochastic differential equation (1.1.2) and suppose that \( b \) and \( \sigma \) are smooth (i.e in \( C^\infty(\mathbb{R}^d) \)) with bounded derivatives vector and matrix fields. Let us introduce the operators
\[
X_0 := \sum_{j=1}^d b_j \partial_{x_j}, \quad \text{and} \quad X_k := \sum_{j=1}^d \sigma^{ij} \partial_{x_j}, \quad k = 1, \ldots, d.
\]

Suppose further that the following condition (Hörmander condition) is satisfied: for any \( x \in \mathbb{R}^d \), \( \mathbb{R}^d = \text{span}(X_1(x), \ldots, X_d(x), [X_{j_1}(x), X_{j_2}(x)], [[X_{j_1}(x), X_{j_2}(x)], X_{j_3}(x)], \ldots |0 \leq j_1, \ldots, j_d \leq d) \), where \( \text{span} \{ \cdot \} \) denotes the vector space generated by any linear combination of the vectors inside the set. Then, for any \( (t, x) \in (0, \infty) \times \mathbb{R}^d \), \( \mathcal{L}^t(X^x) \) has density \( p_t(x, \cdot) \), which viewed as a function over \((0, \infty) \times \mathbb{R}^d \times \mathbb{R}^d \) is smooth.

In particular, strong solutions to SDEs of the form (1.1.2) where the drift is smooth and the noise is additive (i.e drift \( \sigma \) constant) have a smooth transition density.

Next we state a property of the transition densities consequence of the so called flow property of Markovian processes and which is stated as follows: For \( s \geq 1 \) and any \( \tau, x \), \( X^x_{\tau+s} = X^{x_\tau}_{\tau+s} \). It is well known that the flow property holds as a consequence of uniqueness of solutions to SDEs, i.e. Lemma 1.1.19 holds in particular for unique solutions to (1.1.2).

**Lemma 1.1.19 (Chapman–Kolmogorov identity).** Consider a real–valued, time–homogeneous Markov process \((X^x_t)_{s \geq 0}\) and assume that for any \( s, x \), it has a density \( (0, \infty) \times \mathbb{R}^d \ni (s, x') \rightarrow p_s(x, x') \). Suppose moreover that its satisfies that for any \( 0 < t < s \) and any \( x \), \( X^{x_{\tau+s}}_{\tau+s} = X^{x_{\tau+s} \tau}_{\tau+s} \). Then, for any \( s \geq \tau \geq 0 \) and \( x, x'' \in \mathbb{R}^d \):
\[
P_s(x, x'') = \int_{\mathbb{R}^d} p_{s-\tau}(x', x'') p_\tau(x, x') dx'. \tag{1.1.5}
\]

For a complete proof (in the non–autonomous scenario) see Appendix A.1.

### 1.1.4 Semigroup theory

This section, introduces the semigroup unifying notation for analysis of PDEs and SDEs. For a detailed analysis see [10–14].
Definition 1.1.20. A family \((\mathcal{P}_t)_{t \geq 0}\) of linear bounded operators over a Banach space \(E\), \(\mathcal{P}_t : E \to E\) for all \(t \geq 0\), is a semigroup if:

1. \(\mathcal{P}_0 = I\), where \(I\) denotes the identity in \(E\);

2. for all \(t, s \geq 0\) we have \(\mathcal{P}_t \mathcal{P}_s = \mathcal{P}_{t+s}\).

If the map \(\mathbb{R}_+ \ni t \to \mathcal{P}_t f \in B(E)\) is continuous for all \(f \in E\), then the semigroup is said to be strongly continuous. A strongly continuous semigroup of bounded linear operators is also called a \(C_0\)-semigroup. A semigroup is a Markov semigroup if:

- \(\mathcal{P}_t 1 = 1\) for all \(t \in \mathbb{R}_+\) (here 1 is the constant function equal to one);

- it is positivity preserving, i.e. if \(f \geq 0\) then \(\mathcal{P}_t f \geq 0\).

Moreover, the semigroup is called Feller if \(\mathcal{P}_t(C_b(E)) \subseteq C_b(E)\) for all \(t > 0\). And it is strong Feller if \(\mathcal{P}_t(B_b(E)) \subseteq C_b(E)\).

Definition 1.1.21. Given a \(C_0\)-semigroup \((\mathcal{P}_t)\), we define its infinitesimal generator as the operator \(L\) whose action is defined as:

\[
Lf := \lim_{t \to 0^+} \frac{\mathcal{P}_t f - f}{t}, \tag{1.1.6}
\]

for all \(f \in \mathcal{D}(L) := \{f \in E : \text{the limit on the RHS of (1.1.6) exists in } E\}\). If \((\mathcal{P}_t)\) is a strongly continuous Markov semigroup, then the pair \((L, \mathcal{D}(L))\) is the Markov generator of the semigroup.

Next we see how all this relates to stochastic process and in particular to SDEs. Suppose \((X_t)_{t \geq 0}\) is a real valued, continuous time and space–homogeneous Markov process with transition measures given by \(p_t(x, dy) = \mathcal{L}(X_t^x)(dy), \forall (t, x) \in \mathbb{R}_+ \times \mathbb{R}^d\). Then, \(\mathcal{P}_t \phi(x) := \mathbb{E}[\phi(X_t)|X_0 = x] = \int_{\mathbb{R}} \phi(y) p_t(x, dy)\) defines a positivity preserving semigroup. Moreover, every normal Markovian process \(X\) (i.e if the image of bounded, measurable functions through the associated semigroup is again bounded and measurable) has a density and satisfies the Chapman–Kolmogorov identity (1.1.5). In fact, this identity is the reason behind \((\mathcal{P}_t)_{t \geq 0}\) being a semigroup, i.e. satisfying the additive semigroup property in definition 1.1.20. Indeed, for all \(0 \leq s \leq t, \phi \in C_b(\mathbb{R}^d)\) and \(x \in \mathbb{R}^d\) fixed, we have

\[
\mathcal{P}_{t+s} \phi(x) = \int_{\mathbb{R}^d} \phi(x') p_{s+t}(x, x') dx' = \int_{\mathbb{R}^d} \phi(x') \left[ \int_{\mathbb{R}^d} p_t(x, x'') p_s(x'', x') dx'' \right] dx' = \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} \phi(x') p_s(x', x'') dx' \right) p_t(x, x'') dx'' = \int_{\mathbb{R}^d} \left( \mathcal{P}_s \phi(x'') \right) p_t(x, x'') dx'' = \mathcal{P}_t \mathcal{P}_s \phi(x).
\]

Moreover, if given \(x \in \mathbb{R}^d\) the process \((X^x_t)_{t \geq 0}\) satisfies (1.1.2) then the generator is given by Itô’s formula as \(L = b \partial_x + 1/2 \text{tr}(\sigma \sigma^* \partial_x^2)\) and it coincides with the infinitesimal generator mentioned in Section 1.1.3.

1.2 Classical Analysis and Probability theorems

The theory we develop in the subsequent chapters need to be justified based on technical results from classical analysis and probability books. We next list the most relevant ones following [1, 15–18].
Before that, we introduce the notation of the $L^p$ spaces. Given a measure space $(\Omega, \mathcal{F}, \mu)$ and an Euclidean space $(E, B(E), \|\cdot\|_p)$, we say that $f : \Omega \to E$ belongs to $L^p(\Omega, \mathcal{F}, \mu; E)$ if \( \int_\Omega \|f(x)\|_p \mu(dx) < \infty \). Moreover, if there is no confusion, the image space $E$ is omitted. We also simplify notation and do not include the $\sigma$–algebra or the measure if these are $B(\mathbb{R}^d)$ and the Lebesgue measure respectively.

Basic functional inequalities we use are:

- **Young's inequality.** For any $\epsilon, a, b \in \mathbb{R}_+$ and $p > 1, p^* > 1$ such that $\frac{1}{p} + \frac{1}{p^*} = 1$, it holds
  \[ ab \leq \frac{(\epsilon a)^p}{p} + \frac{b^{p^*}}{p^*e^{p^*}}. \]

- **Jensen's inequality.** For any $t \in [0, 1]$, $x, y \in \mathbb{R}^d$ and $f \in B(\mathbb{R}^d)$ convex,
  \[ f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y). \]

  In the context of probability theory: for any $X$ random variable and $f \in B(\mathbb{R}^d)$ convex function,
  \[ f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]. \]

- **Hölder's inequality.** Given any measure $q \in \mathcal{P}(\mathbb{R}^d)$ and functions $f \in L^p(q)$ with $p \in [1, \infty]$ and $g \in L^{p^*}(q)$ with $p^* > 1$ such that $\frac{1}{p} + \frac{1}{p^*} = 1$, the following holds
  \[ \|fg\|_q \leq \|f\|_p\|g\|_{p^*}. \]

  A particularly relevant subcase is provided by $p = p^* = 2$ in which case the inequality receives the name of Cauchy–Schwarz inequality.

A theorem which lies at the bottom of almost every argument involving random variables and which allows us abuse notation and talk indiscriminately about convergence of random variables when we actually mean convergence of measures, is the Skorohod representation theorem. Loosely, this result translates convergence of probability measures (in the weak convergence sense defined below) into convergence of random variable whose distribution coincides with them in certain probability space.

**Definition 1.2.1.** Let $(E, d)$ be a metric space and let $\mathcal{F}$ be a $\sigma$–algebra on $E$ containing the topology generated by the distance $d$. Let $(\mu_n)_{n \in \mathbb{N}}$ be a sequence of positive probability measures defined on $\mathcal{F}$. It is said to converge weakly to $\mu_\infty$ if for any bounded, continuous function $f : E \to \mathbb{R}$ we have

\[ \int_E f(x)\mu_n(dx) \to \int_E f(x)\mu_\infty(dx), \quad \text{as} \quad n \to \infty. \]

**Theorem 1.2.2.** (Skorohod's Representation Theorem [19]) Let $(\mu_n)_{n \in \mathbb{N}}$ be a sequence of probability measures on a metric space $(E, d)$ such that $\mu_n$ converges weakly to some probability measure $\mu_\infty$ on $E$ as $n \to \infty$. Suppose also that the support of $\mu_\infty$ is separable. Then there exist $E$–valued random variable $X_n$ defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that the law of $X_n$ is $\mu_n$ for all $n$ (including $n = \infty$) and such that $(X_n)_{n \in \mathbb{N}}$ converges to $X_\infty$, $\mathbb{P}$–almost surely.
**Definition 1.2.3.** Let \((E, d)\) be a separable metric space, and let \(\mathcal{F}\) be a \(\sigma\)-algebra on \(E\) containing the topology generated by the distance \(d\). Let \(\Pi\) be a collection of (possibly signed or complex) measures defined on \(\mathcal{F}\). The collection \(\Pi\) is called tight (or sometimes uniformly tight) if, for any \(\varepsilon > 0\), there is a compact subset \(K_\varepsilon\) of \(E\) such that, for all measures \(\mu \in \Pi\), \(\mu(K_\varepsilon) > 1 - \varepsilon\).

**Theorem 1.2.4.** (Prokhorov Theorem [20, Theorems 6.1, 6.2]) Let \(\Pi\) be a family of probability measures on a complete, separable metric space \((E, d)\). This family is relatively compact if and only if it is tight.

Next we state the most classical theorem which allows us to prove existence of a solution to an equation: Banach Fixed Point Theorem, for which we need to define the concept of a contraction.

**Definition 1.2.5.** Let \((E, d)\) be a complete metric space (i.e a metric space where every Cauchy sequence is convergent). Then a map \(T : E \to E\) is called a contraction mapping on \(E\) if there exists \(r \in [0,1)\) such that \(d(T(x), T(y)) \leq r d(x, y)\) for all \(x, y \in E\).

**Theorem 1.2.6.** (Banach Fixed Point Theorem [16, Theorem 1.1]) Let \((E, d)\) be a non-empty complete metric space with a contraction mapping \(T : E \to E\). Then \(T\) admits a unique fixed-point \(x^*\) in \(E\) (i.e. \(T(x^*) = x^*\)). Furthermore, \(x^*\) can be found as follows: start with an arbitrary element \(x_0 \in E\) and define a sequence \((x_n)_{n \in \mathbb{N}}\) by \(x_n = T(x_{n-1})\) for \(n \geq 1\). Then \(\lim_{n \to \infty} x_n = x^*\).

Banach Fixed Point Theorem provides indeed existence of solution to the equation \(T(x) - x = 0\) in an iterative manner. The most challenging aspect left for proving existence of certain problem is therefore establishing the appropriate framework, i.e. the space on which the problem lives and the contraction property. A useful theorem used for proving existence of invariant measure (which is the problem we focus on and can be cast into a fixed point problem) is obtaining the contraction inequality.

A crucial point where we need classical functional analysis results is that of exchanging limits.

**Theorem 1.2.7.** (Moore–Osgood Theorem [21, Theorem 1]) Let \(X\) be a metric space, \(E\) one of its subsets and \(x\) a limit point of \(E\). Suppose that \(f_n : X \to \mathbb{R}\) (for each \(n \in \mathbb{N}\)) and \(f : X \to \mathbb{R}\) are functions and \(A_n\) are numbers. If
\[
\lim_{n \to \infty} f_n(x) = f(x), \quad \text{uniformly on } X
\]
and
\[
\lim_{y \to x} f_n(y) = A_n, \quad \text{pointwise over } N,
\]
then the double limit exists and
\[
\lim_{n \to \infty} \lim_{y \to x} f_n(y) = \lim_{y \to x} \lim_{n \to \infty} f_n(y).
\]

Another crucial point where we need classical functional analysis results is that of differentiating under the sign of the integral. The next list is a quick overview of results which can be used to exchange limits.
Theorem 1.2.8. (Leibniz formula [18, Theorem 12.14]) Let $f : \mathbb{R} \times \mathbb{R}_+ \to \mathbb{R}$ be a function such that $f, \partial_x f \in C(\mathbb{R} \times \mathbb{R}_+)$ in some region of the xt–plane, including $a(x) \leq t \leq b(x), x_0 \leq x \leq x_1$. Also suppose that the functions $a, b : \mathbb{R} \to \mathbb{R}$ are both continuous and both have continuous derivatives for $x_0 \leq x \leq x_1$. Then, for $x_0 \leq x \leq x_1$,

$$
\frac{d}{dx} \left( \int_{a(x)}^{b(x)} f(x, t) \, dt \right) = f(x, b(x)) \cdot \frac{d}{dx} b(x) - f(x, a(x)) \cdot \frac{d}{dx} a(x) + \int_{a(x)}^{b(x)} \frac{\partial}{\partial x} f(x, t) \, dt.
$$

In particular, if the limits of integration are constants and the integrand is continuous, we can take the derivative inside the integral.

Definition 1.2.9. A family $(X_t)_{t \in I} \subseteq L^1(\mathbb{R}^d)$ is said uniformly integrable if $\lim_{x \to \infty} \sup_{i \in I} \mathbb{E}[|X_t| \mid X_t \leq x] = 0$.

Theorem 1.2.10. (De La Vallée Poussin Theorem [22, Theorem 2.4.4]) A family $(X_t)_{t \in I} \subseteq L^1(\mathbb{R}^d)$ is uniformly integrable if and only if there exists $\psi \in \Phi := \{ \psi : \mathbb{R} \to \mathbb{R} \mid \lim_{x \to \infty} \psi(x)/x = +\infty \}$ such that $\sup_{i \in I} \mathbb{E}[\psi(|X_t|)] < \infty$.

Theorem 1.2.11. (Vitali’s Convergence Theorem [17, Chapter 4]) Let $(\Omega, \mathcal{F}, \mu)$ be a measure space with $\mu(\Omega) < \infty$. Let $(f_n) \subset L^p(\Omega, \mathcal{F}, \mu)$ and $f$ be an $\mathcal{F}$–measurable function. Then, the following are equivalent:

- $f \in L^p(\Omega, \mathcal{F}, \mu)$ and $(f_n)$ converges to $f$ in $L^p(\Omega, \mathcal{F}, \mu)$;
- The sequence of functions $(f_n)$ converges in $\mu$–measure to $f$ and $(|f_n|^p)_{n \geq 1}$ is uniformly integrable.

1.3 Kolmogorov equations

Forward Kolmogorov or Fokker–Planck and backward Kolmogorov PDEs are englobed under the term Kolmogorov equations. They lie at the interface between PDEs and SDEs and have been studied intensively during the last century. In a nutshell, they are PDEs representing the dynamics of the associated semigroup and transition density for the solution to a SDE with good enough (we will quantify this statement later on this section) coefficients. Kolmogorov for time-continuous and differentiable Markov processes on a finite, discrete state space; and afterwards Feller for more general state spaces, derived independently these equations under quite natural conditions. This section is devoted to their informal derivation within both the probabilistic and semigroup approach. We moreover present the most relevant PDE results which are used to study the well–posedness of their solution.

1.3.1 Derivation

The next discussion is based on material from [5, 9, 11, 23].

Let $(X_t^x)_{t \geq 0}$ denote a time–homogeneous Markov process with transition density $p$, i.e. $\mathcal{L}(X_t^x)(dy) = p_t(x, y) \, dy, \forall t, x, y$. The time homogeneous Markov process can be described through the semigroup formed by the operators $(\mathcal{P}_t \phi)(x) = \mathbb{E}[\phi(X_t^x)]$, whose additive semigroup property relies on the Chapman–Kolmogorov identity. Moreover, if we define the
function \( \mathbb{R}_+ \times \mathbb{R}^d \ni (t,x) \mapsto V(t,x) = \mathcal{P}_t \phi(x) \) for a given \( \phi \in C_b(\mathbb{R}^d) \), then we can calculate its time derivative using its infinitesimal generator \( L \). Formally, by the Definition 1.1.21 we have for any \( x \in \mathbb{R}^d \):

\[
L \phi(x) = \lim_{t \to 0^+} \frac{\mathcal{P}_t \phi(x) - \mathcal{P}_0 \phi(x)}{t} = \partial_t (\mathcal{P}_t \phi(x))|_{t=0} = \partial_t V(0,x).
\]

Now, due to the semigroup property, one can extend the above identity for all times and obtain that in fact \( LV(t,x) = \partial_t V(t,x) \). Moreover, at time \( t = 0 \), from the definition of \( (\mathcal{P}_t)_{t \geq 0} \), we have \( V(0,x) = \phi(x) \) and obtain that the following **Kolmogorov PDE** must be satisfied:

\[
\partial_t V(t,x) = LV(t,x); \quad V(0,x) = \phi(x); \quad \forall (t,x) \in (0,\infty) \times \mathbb{R}^d. \tag{1.3.1}
\]

This is an analysis based approach. A probabilistic approach involves applying Itô’s formula to the process \( \{\phi(X^x_s)\}_{s \geq 0} \) and taking expectations in order to obtain (after a stopping time argument justifying that the process \( \left( \int_0^t \partial_x \phi(X^x_s) \sigma(X^x_s) dB_s \right)_{t \geq 0} \) is indeed a martingale):

\[
d\mathbb{E}[\phi(X^x_t)] = \mathbb{E}[L \phi(X^x_t)] ds.
\]

This approach is repeated a few times in full detail throughout the document (e.g. Lemma B.1.3) and is not repeated here. This once again leads to the equation \( LV(t,x) = \partial_t V(t,x) \) being satisfied.

The specific statement of the first Kolmogorov theorem and its rigorous proof can be found for the one dimensional case in [9, Theorem 2.1]. We state it below in its original shape, which is one time–inversion away from formula (1.3.1). The \( d \)–dimensional case, together with the regularity and growth of the solution are covered in [5, Lemma 5.10].

**Theorem 1.3.1.** [9, Theorem 2.1] Let \( \phi(x) \in C_b(\mathbb{R}^d) \) and for \( X \) solving (1.1.3) and fixed \( 0 \leq s \leq t \), define the function \( u(s,x) := \mathbb{E}[f(X^x_s)] \). Furthermore, assume that \( b, \sigma \) are smooth in both components. Then,

\[
\partial_s u = \sum_{i=1}^d b_i(s,x) \partial_{x_i} u + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^m (\sigma \sigma^\top)^{ij}(s,x) \partial_{x_i}^2 u; \quad u(t,x) = \phi(x), \quad \forall s \in [0,t].
\]

The reciprocal holds and is known as the Feynman–Kac formula. An informal derivation can be found on pp. 68 in [9].

Now, for every \( t \geq 0 \), \( \mathcal{P}_t \) acts on bounded, continuous functions, whereas its adjoint semigroup \( (\mathcal{P}_t^*)_{t \geq 0} \) acts on probability measures. Indeed, for all \( H \in B(\mathbb{R}^d), \mu \in \mathcal{P}(\mathbb{R}^d), t \geq 0 \) the action of the adjoint semigroup \( (\mathcal{P}_t^*)_{t \geq 0} \) is defined as follows:

\[
\mathcal{P}_t^* \mu(H) = \int_{\mathbb{R}^d} \mathbb{P}(X_t \in H|X_0 = x) d\mu(x) = \int_{\mathbb{R}^d} p_t(x,H) d\mu(x);
\]

since for every \( \phi \in C_0^\infty(\mathbb{R}^d), \mu \in \mathcal{P}(\mathbb{R}^d) \) they are related in the following way:

\[
\int_{\mathbb{R}^d} \mathcal{P}_t \phi(x) d\mu(x) = \int_{\mathbb{R}^d} \phi(x) d(\mathcal{P}_t^* \mu)(x). \tag{1.3.2}
\]
Therefore, if $L^*$ is the adjoint operator of $L$ (i.e defined by $C^2(\mathbb{R}^d) \ni f \mapsto L^*f = \partial_x(f b) - \frac{1}{2} \text{tr} (\sigma \sigma^*)$) if the process $X$ is given by (1.1.2) for any test function $h$, (1.3.2) implies that the following must be satisfied:

$$\int_{\mathbb{R}^d} L\phi h \, dx = \int_{\mathbb{R}^d} \phi L^* h \, dx.$$  

And so, if the process $X$ has a transition density, on one hand $p_0(x, x') = \delta_x(x')$ for all choice of parameters and on the other hand (after integrating by parts the above equality): for fixed $x \in \mathbb{R}^d$,

$$\partial_t p_t(x, x') = -L^* p_t(x, x'), \quad \forall (t, x') \in \mathbb{R}_+ \times \mathbb{R}^d.$$  

In other words, the transition density of the process $X$ must satisfy the Fokker–Planck or forward Kolmogorov equation:

$$\partial_t p_t(x, x') = -L^* p_t(x, x'); \quad p_0(x, x') = \delta_x(x'); \quad \forall (t, x') \in \mathbb{R}_+ \times \mathbb{R}^d. \quad (1.3.3)$$

The probabilistic derivation works in a similar fashion: given any $h \in C_0^\infty(\mathbb{R}^d)$ and $t \in \mathbb{R}_+$, we have that backward Kolmogorov equation implies (since Leibniz formula applies due to continuity):

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h(x') \partial_x p_s(x, x') \, dx' \, dx = \int_{\mathbb{R}^d} \partial_s (\mathbb{E}[h(X_s^x)]) \, dx = \int_{\mathbb{R}^d} \partial_s (\mathbb{E}[h(X_s^x)] - h(x)) \, dx = \int_{\mathbb{R}^d} LV(s, x) \, dx.$$  

Next we reduce our analysis to processes given by SDEs of the form (1.1.2) in order to provide a more detailed derivation. We substitute the definition of $L$ and integrate by parts to further develop the above (since Fubini applies due to continuity):

$$\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h(x') \partial_x p_s(x, x') \, dx' \, dx = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \left( b(s, x) \partial_x V(s, x) + \frac{1}{2} \text{tr} (\sigma \sigma^*) (s, x) \partial_x^2 V(s, x) \right) \, dx \, dx'$$

$$= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} b(s, x) \partial_x (h(x') p_s(x, x')) + \frac{1}{2} \text{tr} (\sigma \sigma^*) (s, x) \partial_x^2 (h(x') p_s(x, x')) \, dx' \, dx$$

$$= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} -h(x') \partial_x (b(s, x) p_s(x, x')) + \frac{1}{2} h(x') \text{tr} (\sigma \sigma^*) (s, x) p_s(x, x')) \, dx' \, dx$$

$$= -\int_{\mathbb{R}^d} \int_{\mathbb{R}^d} h(x') L^* (p_s(x, x')) \, dx' \, dx.$$  

In other words, the Fokker–Planck equation we obtained previously using semigroup methodology, is obtained again in the weak sense. If there exists a strong solution, consequently it is also a weak solution.

The specific statement of the second Kolmogorov theorem and its rigorous proof can be found for the one dimensional case in [9, Theorem 2.2]. The $d$–dimensional case is covered in [8, Theorem 6.4.7].

**Theorem 1.3.2.** [9, Theorem 2.2] Let $X$ be a Markov process which solves (1.1.2) and suppose that it has a smooth transition density (i.e for each fixed $s \geq 0$ and $(t, x) \in \mathbb{R}_+ \times \mathbb{R}^d$, there
exists \( p \) such that \( \mathcal{L}(X_t^{x,x})(dy) = p_t^x(x,y)dy \). Assume moreover that \( b, \sigma \) are smooth in both components. Then,

\[
\partial_t p_t^x = -\sum_{i=1}^d \partial_y \left( b_i(t,y) p_t^x \right) + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^m \partial^2_{y_i y_j} \left( (\sigma \sigma^*)^{ij}(t,y) p_t^x \right); \quad p_t^x(x,y) = \delta_0(x-y), \quad \forall (t,y).
\]

A possible link between the forward and backward equations is provided by the following argument which we present in the non–autonomous case. Consider, for regular enough \( \phi \), we consider the backward Kolmogorov function \( V(s,x) = \mathbb{E}[\phi(X_T^{x,x})] \). We next define the function \([0,(1−\tau)] \times \mathbb{R}^d \ni (t,y) \mapsto v(t,x) := V(1−t,x)\), which satisfies (as a consequence of \( V \) satisfying the backward Kolmogorov equation (1.3.1)):

\[
\begin{cases}
\partial_t v(t,x) + L(t,x)v(t,x) = 0, & (t,x) \in [0,1−\tau) \times \mathbb{R}^d \\
v(1−\tau,x) = \phi(x), & x \in \mathbb{R}^d;
\end{cases}
\]  

(1.3.5)

where \( L \) is the infinitesimal generator associated to the SDE (1.3.4). Notice that due to the equation (1.3.5) (first used to simplify Itô’s formula expression and second for substituting the terminal condition), for irregular terminal condition \( x \mapsto \phi(x) = \delta_x(x') \) with \( x' \in \mathbb{R}^d \) fixed, we have for all \( (t,y) \in [0,1−\tau] \times \mathbb{R}^d \):

\[
v(t,x) = \mathbb{E}[v(1−\tau,X_{1−\tau}^t,x')] = \mathbb{E}[\delta_x(X_{1−\tau}^t,x')] = \int_{\mathbb{R}^d} \delta_x(z)p_{1−\tau}^t(x,z)dz = p_{1−\tau}^t(x,x').
\]  

(1.3.6)

This means that our object of interest \( p_t^x(x,x'') \), transition density for the process \( X \) going from \((\tau,x)\) to \((t,x'')\), satisfies on one hand

- for \((\tau,x)\) fixed

\[
\partial_t p_t^x(x,x'') - L^*(t,x'')p_t^x(x,x'') = 0, \quad (t,x'') \in (\tau,1) \times \mathbb{R}^d;
\]

\[
p_t^x(x,x'') = \delta_x(x''),
\]

which is the usual forward Kolmogorov equation satisfied by the density of a stochastic process \((X_t^{x,x})_{t \in (\tau,1)}\), and on the other hand,

- for \((\tau,x'')\) fixed

\[
\partial_t p_{1−\tau}^t(x,x'') + L(t,x)p_{1−\tau}^t(x,x'') = 0, \quad (t,x) \in (0,1−\tau) \times \mathbb{R}^d;
\]

\[
p_{1−\tau}^t(x,x'') = \delta_x(x),
\]  

(1.3.7)

the backward Kolmogorov equation associated to the same process \((X_{t}^{x,x})_{t \in (\tau,1)}\).
1.3.2 Results from PDEs literature

One of the benefits of this link between SDEs and PDEs provided by Kolmogorov equations is the fact that PDEs field, being an older area of study, has a bigger set of results which can be adopted and whose implications to the SDE field one can study. This section presents results in the PDE literature which have the most impact on the regularity and properties of solutions to SDEs. For general parabolic PDE results, see [2], [24] and [24].

First let us establish the notation: given some coefficients, $b$ and $\sigma$, we say that the second order operator $L = b \partial_x + 1/2 \text{tr} (\Sigma \Sigma^* \partial_x^2)$ is uniformly elliptic over $[0, T] \times \mathbb{R}^d$ if there exists $\kappa > 0$ such that for all $(t, x, \xi) \in [0, T] \times \mathbb{R}^d$, we have $\xi^* \Sigma \Sigma^* (t, x) \xi \geq \kappa |\xi|^2$. If $L$ is uniformly elliptic, then $\partial_t - L$ is said to be uniformly parabolic and the equation $(\partial_t - L)(f)(t, x) = 0$ (for $f \in C^2([0, T] \times \mathbb{R}^d)$) is a parabolic PDE.

We start by listing results which can be adapted to provide regularity and estimates of the transition density and of the solution to the backward Kolmogorov equation based on the parametrix technique.

**Theorem 1.3.3.** [25, Theorem 9.4.2, Remark below display (9.4.18)] Suppose that the generator $L = b \partial_x + 1/2 \text{tr} (\Sigma \Sigma^* \partial_x^2)$ is uniformly elliptic on $[0, T] \times \mathbb{R}^d$ and that the coefficients $b, \sigma \in C_b([0, T] \times \mathbb{R}^d)$ and are Hölder uniformly continuous in $x$ in bounded subsets of $[0, T] \times \mathbb{R}^d$. Suppose moreover that the principal coefficients are continuous in $t$ uniformly with respect to $x \in \mathbb{R}^d$. Then for any $\tau, x \in [0, T] \times \mathbb{R}^d$ there exists a solution $(t, x') \mapsto \Gamma_t^\tau (x, x')$ to

$$\partial_t \Gamma_t^\tau (x, x') = L(t, x) \Gamma_t^\tau (x, x'), \quad \Gamma_0^\tau (x, x') = \delta_x (x'), \quad \forall (t, x') \in [\tau, T] \times \mathbb{R}^d;$$

called the fundamental solution, such that

$$|\partial_x^m \Gamma_t^\tau (x, x')| \leq \frac{C}{(t - \tau)^{(d+|m|)/2}} \exp \left(-c \frac{|x - x'|^2}{t - \tau}\right),$$

for $0 \leq |m| \leq 2$, where $C, c > 0$ are some positive constants.

**Theorem 1.3.4.** [26, Theorem VI.2 and Theorem VI.5] Suppose that the generator $L = b \partial_x + 1/2 \text{tr} (\sigma \Sigma^* \partial_x^2)$ is uniformly elliptic on $[0, T] \times \mathbb{R}^d$. Assume that the diffusion $\sigma \in C^{0,2} ([0, T] \times \mathbb{R}^d)$, is continuous in $t$ uniformly in $x$. For the drift we assume $b \in C^{0,2} ([0, T] \times \mathbb{R}^d)$. Additionally we assume that there exist $M_0 \geq 0$ and $0 < \lambda < 1$ such that for all $t \in [0, T]$ and $x, y \in \mathbb{R}^d$,

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq M_0 |x - y|^\lambda;$$

and there exist $M, \epsilon > 0$ such that for all $(t, x) \in [0, T] \times \mathbb{R}^d$,

$$|b(t, x)| \leq M(1 + |x|), \quad |\partial_x b(t, x)| \leq M(1 + |x|)^{1-\epsilon}, \quad |\partial_x^2 b(t, x)| \leq M(1 + |x|)^{3-\epsilon};$$

and

$$|\sigma(t, x)| \leq M, \quad |\partial_x \sigma(t, x)| \leq M(1 + |x|)^{1-\epsilon}, \quad |\partial_x^2 \sigma(t, x)| \leq M(1 + |x|)^{2-\epsilon}.$$

Then for any $\tau, x \in [0, T] \times \mathbb{R}^d$ there exists a solution, unique in the class of bounded functions, for all $(t, x') \in [\tau, T] \times \mathbb{R}^d$ to:

$$\partial_t Z_t^\tau (x, x') = L(t, x) Z_t^\tau (x, x'), \quad (1.3.8)$$

$$Z_0^\tau (x, x') = \delta_x (x'), \quad (1.3.9)$$
so called fundamental solution, such that

$$|\partial^m_x Z^*_t(x, x')| \leq C_m \sum_{k=1}^{\lfloor m \rfloor} (t-\tau)^{-(d+\lfloor m \rfloor-k)/2} (1+|x|^k) \exp(-c|x-x'|^2 + \eta|x|^q - \eta|x'|^q),$$

for $0 \leq |m| \leq 2$, where $C_m, c > 0, \eta \geq 0$ are some constants. Moreover, the solution is unique within the family of bounded functions. Additionally, the solution to the dual of (1.3.8):

$$\partial_\tau Z^*_t(x, x') = L^*(t, x) Z^*_t(x, x')$$  (1.3.10)

exists. Its fundamental solution, $\hat{Z}^*_t(x, x')$, moreover satisfies the normality property: $Z^*_t(x, x')$ when regarded as a function of $(x, t)$ is the fundamental solution of the Cauchy problem (1.3.8), whereas its Hermitian–conjugate matrix $\hat{Z}^*_t(x, x')$, when regarded as a function of $(x', t)$ is the fundamental solution of the adjoint Cauchy problem:

$$\partial_\tau \hat{Z}^*_t(x', x) = L^*(t, x') \hat{Z}^*_t(x', x); \quad \hat{Z}^*_t(x', x) = \delta_{x'}(x)$$  (1.3.11)

In other words,

$$\hat{Z}^*_t(x, x') = \hat{Z}^*_t(x', x).$$

Based on these primary results, other were developed at the interface of PDEs and SDEs. Already in the framework of Kolmogorov equations, and relying heavily on the regularity and growth of the coefficients we have the next two theorems. We state them here and compare them to our results in Chapter 2. First, estimates for the solution to the backwards Kolmogorov equation were provided by Talay for smooth and time–homogeneous coefficients.

**Theorem 1.3.5.**  [27, Theorem 3.4] Consider the SDE (1.1.2) and assume that:

- $b \in C^\infty(\mathbb{R}^d)$ with bounded derivatives of any order;
- $\sigma \in C^\infty(\mathbb{R}^d)$ bounded with bounded derivatives of any order;
- such that there exist $M > 0$ and a compact set $K$ such that $\langle x, b(x) \rangle \leq -M|x|^2, \forall x \in \mathbb{R}^d - K$;
- assume that the operator $L = b(x)\partial_x + \frac{1}{2} \text{tr}(\sigma\sigma^*(x)\partial_x^2)$ is uniformly elliptic;
- $f \in C^\infty$ with at most polynomial growth at infinity.

Then, for any multi-index $I \in \mathbb{N}^d$, there exist positive constants $\Gamma_I, \gamma_I$ and an integer $s_I$ such that the function $V(t, x) = \mathbb{E}[f(X_t^x)]$ satisfies

$$|\partial^I_x V| \leq \Gamma_I (1 + |x|^{s_I}) e^{\gamma_I t}.$$  

We next present a result by Pardoux and Veretennikov on regularity of transition density depending on a parameter. This is needed for example when decoupling a McKean–Vlasov process in such a way that we obtain an auxiliary non–autonomous SDE in Chapter 3.

**Theorem 1.3.6.**  [28, Proposition 2 and 3] Given coefficients $b, \sigma$ and values $x, y \in \mathbb{R}^d$, consider an SDE of the form

$$dX_t^y = b(X_t^y, y) dt + \sigma(X_t^y, y) dB_t, \quad t \geq 0;$$  (1.3.12)
and the associated infinitesimal generator $L(x, y) = \sum_{i,j=1}^{d}(\sigma \sigma^*(x, y))^{ij}\partial^2_{x_ix_j} + \sum_{i=1}^{d} b_i \partial_{x_i}$. Assume that there exist $\lambda, \Lambda > 0$ such that $\lambda I_d \leq (\sigma \sigma^*(x, y)) \leq \Lambda I_d$, $b$ is bounded and both $b$ and $\sigma$ are Hölder continuous in $x$ uniformly in $y$. Then, the transition density $p_t(x, x'; y)$ exists and satisfies that for any $T > 0$ there exist some $C, c > 0$ such that for any $0 \leq |m| \leq 2$ and $0 \leq t \leq T$, 

$$|\partial^m_x p_t(x, x'; y)| \leq Ct^{-(d+|m|)/2} \exp(-c|x-x'|^2/2).$$

Additional derivatives are obtained at the expense on assuming that the coefficients have higher order bounded derivatives. Suppose moreover that \( \lim_{|x| \to \infty} \sup_y (x, b(x, y)) = -\infty \), that for each $y \in \mathbb{R}^d$, $b(\cdot, y), \sigma(\cdot, y) \in C^1_b(\mathbb{R}^d)$ for some $n \in \mathbb{N}$ and with uniform bounds in $y$. Then, for any $k, j \in \mathbb{R}_+$ there exist $C, m > 0$ such that for all $x, x', y \in \mathbb{R}^d$, $t \geq 1$:

$$|\partial^m_{x'} p_t(x, x'; y)| \leq C \frac{1 + |x|^m}{1 + |x'|^j}.$$

Moreover, provided that $n \geq 1$ and for each $y \in \mathbb{R}^d$, $b(\cdot, y), \sigma(\cdot, y) \in C^{(n-2)+}_{b}(\mathbb{R}^d)$ (again with uniform in $y$ bounds), we have

$$|\partial^m_{x'} p_t(x, x'; y)| \leq C \frac{1 + |x|^m}{(1 + t)^k(1 + |x'|^j)}, \quad \forall x, x' \in \mathbb{R}^d, t \geq 1.$$

A particularly surprising application of the theory of PDEs to the SDEs field has been presented in [29, Theorem 1.2], where the authors unveil the phenomenon of loss of regularity in Kolmogorov PDEs. For first order Kolmogorov PDEs with smooth coefficients, regularity preservation of solutions to (1.3.1) is a well known result (in other words, if $\sigma(x) = 0$ for all $x \in \mathbb{R}^d$ and if the initial function $\phi : \mathbb{R}^d \to \mathbb{R}^d$ in (1.3.1) is smooth, there exists a unique smooth classical solution of (1.3.1)). For more information see [29]. Furthermore, for second order Kolmogorov PDEs (i.e $\sigma \neq 0$), we even observe a smoothing effect. This is indeed the case of operators falling under the Hörmander Theorem's 1.1.18 hypothesis. However, there is a regime where [29, Theorem 1.2] reveals that even though we assume $\phi \in C^\infty(\mathbb{R}^d)$, the solution to (1.3.1) $V(t, \cdot)$ does not belong to $C^1(\mathbb{R}^d), \forall t > 0$. In probabilistic terms this example translates as: smooth functions with compact support may be mapped to non-smooth functions by the transition semigroup associated to an SDE.

### 1.3.3 Ergodicity

This section is devoted to the particular behaviour in time of transition densities. More explicitly, we present the concepts of invariant measure and ergodic behaviour of a Markovian process.

**Definition 1.3.7.** Let $(\mathcal{P}_t)_{t \geq 0}$ be a Markov semigroup over the $\mathbb{R}^d$ Euclidean space. A probability measure $\mu \in \mathcal{P}(\mathbb{R}^d)$ is said to be invariant (or stationary) for $(\mathcal{P}_t)_{t \geq 0}$ if for any $\phi \in B_b(\mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} (\mathcal{P}_t \phi)(x) \mu(dx) = \int_{\mathbb{R}^d} \phi(x) \mu(dx), \forall t \geq 0.$$

Given a time–homogeneous $\mathbb{R}^d$–valued Markov process, we say that a measure $\mu$ is invariant for $X$ if it is invariant for the associated Markov semigroup.
Informally, if a stochastic process hits at any time one of its invariant measures, then it “sticks” to it in the sense that, for any future time, the stochastic process will always be distributed as its invariant measure. Time series analysis must check such a stationarity condition in order to access linear regression tools.

**Definition 1.3.8.** A time-homogeneous (continuous time) Markov process is ergodic if the associated semigroup admits a unique invariant measure.

In terms of applications, such a property is desirable because it translates to equilibrium states for the system modelled by the Markovian process. Moreover, it is shown in [10, Theorem 3.2.4] that, for $C_0$–Markov semigroups, ergodicity is equivalent to the existence of a measure $\mu$ such that for any $\psi \in L^2(\mathcal{D}(L), \mu),$

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t (PS\psi)(x) \, ds = \int_D \psi(x) \mu(dx), \quad \text{in } L^2(\mathcal{D}(L), \mu).$$

In other words, time averaging can be used as a proxy for samples of equilibrium states.

Again, let us relate the existence of such an invariant measure to the theory of PDEs. Note that, loosely speaking, one can take time derivatives and obtain the following equivalence for any strongly continuous semigroup in $L^p(\mathbb{R}^d, \mu),$ with generator $L$:

$$\int_{\mathbb{R}^d} (P_t\phi)(x) \mu(dx) = \int_{\mathbb{R}^d} \phi(x) \mu(dx), \quad \forall t \geq 0 \iff \int_{\mathbb{R}^d} L\phi d\mu = 0, \quad \forall \phi \in \mathcal{D}(L).$$

Now, if we want to close the circle and characterize this property in terms of the generator $L^*$ of $P^*$, the problem is more delicate. This is because the adjoint operator might not be strongly continuous in general. However, if the space on which the operators are defined is reflexive, then strong continuity of the semigroup implies strong continuity of the adjoint semigroup. So, formally and under appropriate assumptions, if we consider the flat $L^2$–adjoint of the generator of $L$, denoted by $L^*$, a measure $\mu$ is invariant for the semigroup if $L^* \mu = 0$ (see [14, Section 1.10]).

There are many ways of setting up assumptions in order to obtain an invariant measure. For each, we include a single reference and the reader is remitted to [30] for a more detailed list of references for:

(i) The approach based on Harris theorem or the Meyn–Tweedie approach with Lyapunov functions [31];

(ii) The approach based on entropy estimates and Poincaré and Sobolev inequalities [32];

(iii) The probabilistic approach based on coupling [33].

We now present two classic theorems based on the approach (i) above.

**Theorem 1.3.9.** (*Harris ergodic theorem [31, Theorem 1.3]*) Let $\mathcal{P}(\cdot, \cdot)$ be a Markov transition kernel defined on a measurable set $(X, \mathcal{F})$, i.e. for each $x \in X$, the function $B \mapsto \mathcal{P}(x, B)$ is a probability measure on $\mathcal{F}$ and for each $B \in \mathcal{F}$, $x \mapsto \mathcal{P}(x, B)$ is $\mathcal{F}$–measurable. The transition kernel defines operators on functions and measures by setting

$$\mathcal{P} f(x) = \int_X f(y) \mathcal{P}(x, dy),$$

...
\[ \mathcal{P}\sigma(B) = \int_X \mathcal{P}(x,B)\sigma(dx). \]

Let us assume that:

1. There exists a function \( U : X \to \mathbb{R}_+ \) and numbers \( \delta \in (0,1) \) and \( K \) such that \( \mathcal{P}U(x) \leq \delta U(x) + K, \forall x \in X; \)

2. There exists a number \( k \in (0,1) \) and a probability measure \( \sigma \) such that \( \sup_{x : U(x) \leq R} \mathcal{P}(x,\cdot) \geq q\sigma(\cdot) \) for some \( R > 2K/(1 - \delta). \)

Then, there exist numbers \( \beta > 0, \hat{\beta} \in (0,1) \) such that, for every pair of probability measures \( \mu_1, \mu_2 \) on \( X, \)

\[ \|\mathcal{P}\mu_1 - \mathcal{P}\mu_2\|_{\beta U} \leq \hat{\beta}\|\mu_1 - \mu_2\|_{\beta U}, \]

where \( \|\mu_1 - \mu_2\|_{\beta U} := \int_{\mathbb{R}^d} (1 + \beta U(x))|\mu_1 - \mu_2|(dx). \)

**Theorem 1.3.10.** [34, Theorem 11] Assume that the SDE (1.1.2) has a unique strong solution and that there exists \( W \in C^2 \) a non-negative function such that \( LW(x) \to -\infty \) as \( |x| \to \infty \), where \( L \) is the associated generator. Then, there exists a unique stationary solution to (1.1.2).

Throughout Chapters 2, 3 and 4, we present a similar approach for obtaining the existence of an invariant measure based on either Lyapunov functions (see Lemma 2.3.12) or monotonicity conditions (Lemma 3.2.10).

### 1.4 Measure space calculus

Let us first establish the framework. Let \((\Omega, \mathcal{F}, \mathbb{P})\), \( \Omega \) Polish space and \( \mathbb{P} \) atomless, be a probability space and \((E, d)\) a metric space. Denote by \( \mathcal{P}(E) \) the probability measures on \( E \).

For any \( p \geq 1 \) and \( \mu, \nu \in \mathcal{P}(E) \), the \( p \)-Wasserstein distance \( W_p(\mu, \nu) \) is stated in Definition 1.1.5.

Moreover, for \((E, d)\) being \( \mathbb{R}^d \) with the Euclidean distance and given a positive, increasing function \( W : \mathbb{R}^d \to \mathbb{R}^+ \), we define the weighted Wasserstein semi–distance as

\[ W_W(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \left[ \int_{E \times E} W(x-y) d\pi(x, y) \right]. \]

Machine learning and optimal transport theory make use of functions of the form \( U : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R} \) applied to the law of a stochastic process. Notice that here the original metric space is the Euclidean space \( \mathbb{R}^d, \), \( d \geq 1, \) and the space of measures defined on it with finite second moment, \( \mathcal{P}_2(\mathbb{R}^d), \) is dotted with the 2-Wasserstein metric. As a consequence, there have been defined many notions of differentiability on measure spaces. In this document we focus mainly on the \( L \)-derivative, but for the benefit of the reader, since we refer to documents considering the linear functional derivative, we define it as well.

The choice of the \( L \)-derivative is motivated by the fact that its definition is based on the lifting of the above type of functions on probability measures \( (U) \) to functions \( \hat{U} : L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d) \to \mathbb{R} \) defined through the relation \( \hat{U}(X) := U(L(X)) \), for any random variable \( X \) belonging to the Hilbert space \( L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d). \) In particular, it provides insight into infinitesimal perturbations of probability measures induced by infinitesimal variations in a linear space of random variables. For Itô processes this is specially interesting because it
alows the study of the sensibility of the law of the solution to changes in datum. For more detail see [35, Vol. 1, Chapter 5, Section 2]. This is a well defined notion due to the fact that under the assumptions that \( \Omega \) is a Polish space and \( \mathbb{P} \) is atomless, for any probability measure \( \mu \in \mathcal{P}_2(\mathbb{R}^d) \), there exists a random variable \( X \in L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d) \) such that \( \mu = \mathcal{L}(X) \) (see [36, Proposition 9.1.11] or [37, Theorem 13.1.1]).

**Definition 1.4.1 (L-derivative).** A function \( U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d \) is said to be L–differentiable at \( \mu_0 \in \mathcal{P}_2(\mathbb{R}^d) \) if there exists a \( \mathbb{R}^d \)–valued random variable \( X_0 \) satisfying \( \mu_0 = \mathcal{L}(X_0) \), such that the lifted version \( \tilde{U} \) of the function \( U \) is Fréchet differentiable at \( X_0 \).

The Fréchet derivative \( D\tilde{U}(X_0) \) is called the representation of the L–derivative of \( U \) at \( \mu_0 \) along the variable \( X \).

Consider a continuously L–differentiable function \( U \) on \( \mathcal{P}_2(\mathbb{R}^d) \), meaning that its lifted version \( \tilde{U} \) has a Fréchet derivative \( X \mapsto D\tilde{U} \) which is a continuous function from \( L^2(\Omega, \mathcal{F}, \mathbb{P}; \mathbb{R}^d) \) into itself. By [35, Proposition 5.25], we know that for any \( \mu_0 \in \mathcal{P}_2(\mathbb{R}^d) \), \( D\tilde{U}(X_0) = \xi(X_0) \) for a deterministic measurable function \( \xi : \mathbb{R}^d \to \mathbb{R}^d \), which is uniquely defined \( \mu_0 \)–almost everywhere on \( \mathbb{R}^d \). The equivalence class of \( \xi \) in \( L^2(\mathbb{R}^d, B(\mathbb{R}^d), \mu_0; \mathbb{R}^d) \), being uniquely defined, is denoted by \( \partial_\mu U(\mu_0) \). We say that \( \partial_\mu U(\mu_0) \) is the L–derivative of \( U \) at \( \mu_0 \) and identify it with a function \( \partial_\mu U(\mu_0) : \mathbb{R}^d \to \mathbb{R}^d \).

The concepts of L–continuity and L–differentiability are discussed in more detail in [5, Section 4.4].

**Definition 1.4.2 (Linear functional derivative).** A function \( U : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \) is said to have linear functional derivative if there exists a function

\[
\frac{\delta U}{\delta m} : \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}
\]

continuous with respect to the product topology, \( \mathcal{P}_2(\mathbb{R}^d) \) being equipped with the 2-Wasserstein distance, such that:

- for any bounded subset \( \mathcal{K} \in \mathcal{P}_2(\mathbb{R}^d) \), the function \( x \mapsto \frac{\delta U}{\delta m}(m)(x) \) is at most of quadratic growth in \( x \) uniformly in \( m \); and

- for all \( m, m' \in \mathcal{P}_2(\mathbb{R}^d) \) the following holds:

\[
U(m) - U(m') = \int_0^1 \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(tm' + (1-t)m)(x)d(m' - m)(x)dt.
\]

**Remark 1.4.3.** Note that the above definition is unique up to a constant. To overcome non–uniqueness, the following normalization condition is usually assumed in the literature

\[
\int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(m)(x)dm(x) = 0.
\]

The relation between the two notions of derivative in the space of measures is captured by

\[
\partial_\mu U(\mu)(\cdot) = \partial_x \frac{\delta U}{\delta m}(\mu)(\cdot), \quad \mu \in \mathcal{P}_2(\mathbb{R}^d),
\]
whenever both notions of derivative are well defined, the function $\frac{\partial U}{\partial m}(\mu) \cdot \text{ is differentiable}$ and the derivative $(\mu, x) \rightarrow \partial_x \left[ \frac{\partial U}{\partial m} \right] (\mu)(x)$ is jointly continuous in $(m, x)$ and is at most of linear growth in $x$, uniformly in $m \in \mathcal{K}$, for any bounded subset $\mathcal{K} \in \mathcal{P}_0(\mathbb{R}^d)$ (see [35]). In a broader context, when the expression $DU(m, x) := \partial_x \left[ \frac{\partial U}{\partial m} \right] (\mu)(x)$ makes sense, it is known as the intrinsic derivative or Lions derivative.

Next we state the generalization of the function spaces introduced in Section 1.1.

**Definition 1.4.4.** A function $U : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ is fully $\mathcal{C}^{1,1}$ if the following mappings

\[
\begin{align*}
\mathcal{P}_2(\mathbb{R}^d) &\ni \mu \mapsto U(\mu) \\
\mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d &\ni (\mu, y) \mapsto \partial_\mu U(\mu, y) \\
\mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d &\ni (\mu, y) \mapsto \partial_y \partial_\mu U(\mu, y)
\end{align*}
\]

are well defined and continuous with respect to the product topology. If additionally $\mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \ni (\mu, y) \mapsto \partial^2_\mu U(\mu, y)$ is well defined and continuous with respect to the product topology, we say instead that $U : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ is fully $\mathcal{C}^{2,1}$.

In general, we say that $U \in \mathcal{C}^{1,2,1}(\{0, \infty\} \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$ if for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ the function $U(\cdot, \cdot, \mu) \in C^{1,2}([0, \infty) \times \mathbb{R}^d)$ and for each $(t, x) \in [0, \infty) \times \mathbb{R}^d$, $\partial_\mu U(t, x, \mu) \in \mathcal{C}^{1,1}(\mathcal{P}_2(\mathbb{R}^d))$. Moreover, all the resulting (partial) derivatives must be jointly continuous in all their components.

Higher order derivatives are defined by induction from Definition 1.4.2 as it follows:

**Definition 1.4.5.** For any $p \geq 1$, the $p$-th order linear derivative of a function $U$, is a continuous function $\frac{\partial^p U}{\partial m^p} : \mathcal{P}_2(\mathbb{R}^d) \times (\mathbb{R}^d)^{p-1} \times \mathbb{R}^d \rightarrow \mathbb{R}$ satisfying:

- for any bounded subset $\mathcal{K} \subset \mathcal{P}_2(\mathbb{R}^d)$, the function $(x, x') \mapsto \frac{\partial^p U}{\partial m^p}(m, x, x')$ is at most of quadratic growth in $(x, x')$ uniformly in $m$; and

- for all $m, m' \in \mathcal{P}_2(\mathbb{R}^d)$ the following holds:

\[
\frac{\delta^{p-1} U}{\delta m^{p-1}}(m', x) - \frac{\delta^{p-1} U}{\delta m^{p-1}}(m, y) = \int_0^1 \int_{\mathbb{R}^d} \frac{\delta^p U}{\delta m^p}(tm' + (1 - t)m)(x, x')d(m' - m)(x')dt,
\]

provided that the $(p - 1)$-th order linear derivative is well defined.

A particularly interesting property is proved in [38]: if for some $p \geq 1$ we have $\partial^p U \in L^\infty(\mathbb{R}^d)$ (i.e bounded uniformly in the space and measure components), then there exists $C_p > 0$ such that

\[
\left| \frac{\partial^p U}{\partial \mu^p}(\mu, y_1, \ldots, y_p) \right| \leq C_p(|y_1|^p + \ldots + |y_p|^p), \quad \forall y = (y_1, \ldots, y_p) \in \mathbb{R}^{d \times p}.
\]

### 1.4.1 McKean–Vlasov SDEs

The theory on measure spaces allows the extension of the theory of SDEs to measure–dependent coefficients. In particular, stochastic processes whose transition functions might depend on the current distribution of the process and not only on the current state, were introduced by McKean in 1966 to model plasma dynamics.
**Definition 1.4.6.** Given $T > 0$, $b : [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ and $\sigma : [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d \times \mathbb{R}^d$ a McKean–Vlasov SDE takes the following form:

$$dX_t^{0, \xi} = b(X_t^{0, \xi}, \mathcal{L}(X_t^{0, \xi})) \, dt + \sigma(X_t^{0, \xi}, \mathcal{L}(X_t^{0, \xi})) \, dB_t, \quad \forall \, t \in [0, T]; \quad X_0^{0, \xi} = \xi. \quad (1.4.1)$$

We say that (1.4.1) has a solution if Definition 1.1.2 is satisfied for (1.4.1) instead of (1.1.2).

Its infinitesimal generator is then given for any $(t, x, \mu) \in [0, \infty) \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$ by $L^\mu = L^\mu(t, x, \mu)$, where for any $U \in \mathcal{C}^{1,2,(1,1)}([0, \infty) \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$ we have:

$$(L^\mu U)(t, x, \mu) = \left( b \partial_x U(t, x, \mu) + 1/2 \text{tr}(\sigma \sigma^* \partial^2_x U)(t, x, \mu) \right)$$

$$+ \int_{\mathbb{R}^d} \left( b(t, x, \mu)(\partial_\mu U(t, x, \mu)(y)) + 1/2 \text{tr}(\sigma \sigma^*(t, x, \mu)\partial_\mu U(t, x, \mu)(y)) \right) \mu(\,d y). \quad (1.4.2)$$

Before stating some available well–posedness results of McKean–Vlasov SDEs, we introduce some notation. Given a domain $D$, we use $\partial D$ for denoting the topological boundary and $\bar{D}$ for the closure. Moreover, we say $(D_k)_{k \in \mathbb{N}}$ is a bounded nested sequence of bounded sub–domains of $D$ if $D_k$ are bounded, $\cup_k D_k = D$ and for all $k \in \mathbb{N}$ we have $\bar{D}_k \subset D_{k+1}$. Finally, we say $\xi$ is $W$–integrable if $\int_{\mathbb{R}^d} W(x) \mathcal{L}(\xi)(\,d x) < \infty$ and is locally integrable if its integral over compact sets is finite.

**Theorem 1.4.7.** (Existence and uniqueness of solution to McKean–Vlasov SDEs [39, Theorem 2.10, Corollary 3.4 and Theorem 4.1]) Let $D \subset \mathbb{R}^d$, $I \subset [0, \infty)$ and assume

1. There exists a so called Lyapunov function: $W \in \mathcal{C}^{1,2,(1,1)}([0, \infty) \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$, $W \geq 0$ and locally integrable random functions $m_1, m_2$ on $I$ such that if $(D_k)_{k \in \mathbb{N}}$ is a bounded nested sequence of bounded sub–domains of $D$, then for all $k \in \mathbb{N}, t \in I, x \in D_k, \mu \in \mathcal{P}(D_k)$, we have,

$$(L^\mu W)(t, x, \mu) \leq m_1(t) W(t, x, \mu) + m_2(t). \quad (1.4.3)$$

2. $\mathcal{L}(\xi)$ can be approximated by a sequence of probability distributions $(\mu^k)_{k}$ such that $\mu^k_0 \Rightarrow \mu$ (weak convergence of probability measures) and for each $k \in \mathbb{N}, \mu^k$ is supported on $D_k$ and for some increasing function $\psi_W : [0, \infty) \to [0, \infty)$ such that $\psi_W(x) \geq x$, we have

$$\int_{D_k} W(0, x, \mu^k) \, d \mu^k(\,d x) \leq \psi_W(W(0, \cdot, \mathcal{L}(\xi))) < \infty. \quad (1.4.4)$$

3. $b, \sigma$ are continuous in the space and measure components in the following sense: if $(\mu_n) \in \mathcal{P}(D)$ satisfy $\sup_n \sup_{t \in I} \int_D W(t, x, \mu^n) \, d \mu^n(\,d x) < \infty$ and if $(x_n \to x, \mu^n \to \mu)$ as $n \to \infty$, then

$$\forall \, t \in I, \quad b(t, x_n, \mu_n) \to b(t, x, \mu) \quad \text{and} \quad \sigma(t, x_n, \mu_n) \to \sigma(t, x, \mu) \quad \text{as} \quad n \to \infty. \quad (1.4.5)$$

4. There exist $c_k > 0$ such that for any $\mu \in \mathcal{P}(D),

$$\sup_{x \in D_k} (|b(t, x, \mu)| + |\sigma(t, x, \mu)|) < c_k \left( 1 + \int_D W(t, y, \mu) \, d y \right). \quad (1.4.6)$$
Then, if there exists a non–negative function $w : I \times D$ such that for all $(k, t, x, \mu) \in \mathbb{N} \times I \times D_k \times \mathscr{P}(D_k)$ we have

$$w(t, x) \leq W(t, x, \mu) \quad \text{and} \quad \inf_{t \in I, x \in \mathbb{R}D_k} w(t, x) \to \infty \text{ as } k \to \infty.$$ \nonumber

and \(\sup_{t \in I} \max(0, M(t)) < \infty\), then there exists a $W$–integrable weak solution to (1.4.1) on $I$. In addition, \(\sup_{t \in I} E[W(t, X_t, \mathcal{L}(X_t))] < \infty\). Moreover, suppose that there exists $\bar{W} \in C^2(\mathbb{R}^d)$ and a locally, non-random functions $g, h$ on $I$ such that for any two solutions $X$ and $Y$ of (1.4.1),

$$\begin{align*}
&\left(b(t, X_t, \mathcal{L}(X_t)) - b(t, Y_t, \mathcal{L}(Y_t))\right)\partial_x(\bar{W})(X_t - Y_t) \\
&+ \frac{1}{2} \text{tr}\left((\sigma(t, X_t, \mathcal{L}(X_t)) - \sigma(t, Y_t, \mathcal{L}(Y_t))) (\sigma(t, X_t, \mathcal{L}(X_t)) - \sigma(t, Y_t, \mathcal{L}(Y_t)))^* \partial^2_x(\bar{W})(X_t - Y_t)\right) \\
&\leq g(t)(\bar{W})(X_t - Y_t) + h(t)\mathcal{W}_r(\mathcal{L}(X_t), \mathcal{L}(X_t)).
\end{align*}$$

Then, the solutions to (1.4.1) are pathwise unique. Finally, if the conditions hold for $I = [0, \infty)$ and $\bar{W}(t, x) \geq |x|^2$ for $x \in D_k$ for some $k \in \mathbb{N}$ and all $t \geq 0$, and if the associated semigroup is Feller, then there exists an invariant measure.

Convergence to invariant measure can be obtained under the Total Variation distance, the Wasserstein distance or even the Weighted Total Variation distance (see [40–43]). In particular, existence and uniqueness of solutions has been proved in [44] for coefficients satisfying Lipschitz and monotonicity conditions. Additionally, they show convergence in $\mathcal{W}_2$ to the stationary solution.

Before finishing the introduction to measure–space calculus, let us mention a few generalizations of the results introduced earlier in the chapter to underlying McKean–Vlasov processes instead of Itô processes. Indeed, given a function $U : \mathscr{D}_2(\mathbb{R}^d) \to \mathbb{R}$ and an SDE (1.4.1), one can differentiate $U(\mathcal{L}(X_t^0, \xi))$ (assuming that $\xi \in L^2$ and there exists a solution to (4.1.1) denoted by $(\bar{X}_t, \bar{b}_t, \bar{\sigma}_t, t \in [0, T])$ using the regular Itô–Wentzell–Lions formula, which can be found in [45, Theorem 3.3]. Let $(\bar{X}_t, \bar{b}_t, \bar{\sigma}_t, t \in [0, T])$ be the independent process copy (i.e., twin) of $(X_t, b_t, \sigma_t, t \in [0, T])$ living in the twin probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$, where we denote expectation by $\bar{E}$. Then, Itô–Wentzell–Lions formula reads as follows: for all $t \in [0, T]$:

$$U(\mathcal{L}(X_t^0, \xi)) = U(\mathcal{L}(\xi)) + \int_0^t \bar{E} \left[\partial_\mu U(\mathcal{L}(X_s^0, \xi), \bar{X}_s^{0, \xi}) \bar{b}(\bar{X}_s^{0, \xi}, \mathcal{L}(\bar{X}_s^{0, \xi}))ight] ds$$

whenever $U$ is fully $\mathcal{C}^2$ and for any compact subset $\mathcal{K} \subset \mathscr{D}_2(\mathbb{R}^d)$,

$$\sup_{m \in \mathcal{K}} \int_{\mathbb{R}^d} \left(\left|\partial_\mu U(m, y)\right|^2 + \left|\partial_y \partial_\mu U(m, y)\right|^2\right) dm(y) < \infty.$$ \nonumber

A more general formula, for random and time dependent functions $U$, can be found in [46, Theorem 3.3].

Moreover, similar to what we have for the regular SDEs, the backward Kolmogorov equation has as analogous known as the Master Equation (see [47, Thorem 7.2] and also [48]): given $U : \mathscr{D}_2(\mathbb{R}^d) \to \mathbb{R}$ and an SDE (1.4.1) with $\mathcal{L}(\xi) = \mu$, the function $[0, T] \times \mathscr{D}_2(\mathbb{R}^d) \ni (t, \mu) \rightarrow$
\[ \mathcal{U}(t, \mu) := U(\mathcal{L}(X^t_i)) \] satisfies:
\[
\begin{cases}
\partial_s \mathcal{U}(s, \mu) - \left[ \int_{\mathbb{R}^d} \partial_\mu \mathcal{U}(s, \mu)(y) b(y, \mu) + \frac{1}{2} \text{tr}(\partial_v \mathcal{U}(s, \mu)(y)) \mu(dy) \right] = 0, & \forall (s, \mu) \in [0, T] \times \mathcal{P}(\mathbb{R}^d); \\
\mathcal{U}(0, \mu) = U(\mu), & \forall \mu \in \mathcal{P}(\mathbb{R}^d).
\end{cases}
\]

(1.4.4)

Notice that in the particular case that \( U(\mathcal{L}(\xi)) = \int_{\mathbb{R}^d} \phi(x) d\mathcal{L}(\xi)(x) \) for some \( \phi : \mathbb{R}^d \to \mathbb{R} \) belonging to \( C^2(\mathbb{R}^d) \), we have \( \frac{\partial U}{\partial \mu}(\mu, x) = \phi(x) \) and \( \partial_\mu U(\mu, x) = \partial_x \phi(x) \). Under these conditions and for deterministic initial data, the Master equation reduces to (1.3.1) (up to a time reversal) where we allow the coefficients to be measure dependent. Chapter 3 is devoted to explicit uniform (in time) estimates for the gradient of the solution to this simpler Master equation, i.e. \((t, x) \mapsto V(t, x) = E[\phi(X^{0, x}_t)] \) for \( X \) solution to (1.4.1).

### 1.5 Particle systems

It is clear why McKean–Vlasov SDEs are a generalization of the SDEs defining the Itô processes. This section addresses the reason behind including dependence of the coefficients on the law of the solution itself being a natural extension. More specifically, this section is devoted to how McKean–Vlasov SDEs arise as limiting (as the number of particles go to infinity) behaviour of interacting particle systems. Since high dimensional particle systems are challenging to analyse, mathematicians study their associated McKean–Vlasov limit and extract pertinent consequences for finite number particle systems.

Before introducing particle systems we define the concept that lies at its bottom: random (\([49, \text{Chapter II. Definition 1.3}]\)) and empirical measures.

**Definition 1.5.1.** Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability measure space equipped with a filtration \((\mathcal{F}_i)_{i \geq 0}\) and \((E, B(E))\) be a Polish space with its Borel \(\sigma\)-field. A random measure on \(\mathbb{R}_+ \times E\) is a family \(\mu = (\mu(\omega; dt, dx) : \omega \in \Omega)\) of non-negative measures (for each \(\omega\) fixed) on \((\mathbb{R}_+ \times E, B(\mathbb{R}_+) \otimes B(E))\) satisfying that for any \(A \in B(\mathbb{R}_+) \otimes B(E)\), \(\mu(\omega; A)\) is \(\mathcal{F}\)-measurable.

A particularly relevant (for us) type of random measures are the empirical measures.

**Definition 1.5.2.** Given a probability space \((\Omega, \mathcal{F}, \mathbb{P})\) and \(N \in \mathbb{N}\) random variables defined on it : \((Y^i_j)_{j=1,...,N}\), the random measure \(\mu^{Y,N} = \frac{1}{N} \sum_{j=1}^{N} \delta_{Y^i_j}\) is called an empirical measure.

Extremely useful, dimension-free, estimates when working with particle systems (e.g. for decomposing the expression of the weak error along the particle system in Theorem 4.4.2) are captured by the following theorem, the proof of which relies on measure space calculus.

**Theorem 1.5.3.** \([50, \text{Theorem 2.12}]\) Let \((\xi_i)_{1 \leq i \leq N}\) be i.i.d. random variables with law \(\mu \in \mathcal{P}_2(\mathbb{R}^d)\) and let \(\mathcal{U} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}\) satisfy the following uniform estimates for \(p = 1, \ldots, 5\): for any \(m \in \mathcal{P}_2(\mathbb{R}^d)\) and \(X_1, \ldots, X_P\) i.i.d random variables distributed as \(m\), there exists a constant \(L(\mathcal{U}, m) > 0\) such that \(E \left[ \sup_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \frac{\delta_{\mathcal{U}}}{\delta \mu^p} (\nu, X_1, \ldots, X_P) \right] \leq L(\mathcal{U}, m)\). Then,

\[
E \left[ \mathcal{U} \left( \frac{1}{N} \sum_{i=1}^{N} \delta_{\xi_i} \right) \right] - \mathcal{U}(\mu) = \mathcal{O} \left( \frac{1}{N} \right),
\]

for the usual \(\Theta\) notation (i.e. the estimate holds up to a positive multiplying constant).
Another key property which makes calculus on measure spaces a valuable tool is due to Lions Proposition on the special behaviour of the empirical measure with respect to the measure derivative. Again we rely on it for decomposing the expression of the weak error along the particle system in Theorems 4.2.3,4.4.2.

**Proposition 1.5.4.** ([45, Proposition 3.1]) Assume that $\mathcal{U} : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ is fully $\mathcal{C}^2$. Then, $\mathbb{R}^d \ni (x^1, \ldots, x^N) \mapsto U(x^1, \ldots, x^N) := \mathcal{U}\left(\frac{1}{N} \sum_{j=1}^N \delta_{x^j}\right)$ belongs to $\mathcal{C}^2(\mathbb{R}^d)$ and the following holds:

\[
\begin{align*}
\delta_{x^j} U(s, x^1, \ldots, x^N) &= \frac{1}{N} \partial_x \mathcal{U}\left(s, \frac{1}{N} \sum_{j=1}^N \delta_{x^j}\right)(x^i) \quad (1.5.1) \\
\text{and} \\
\delta_{x^j}^2 U(s, x^1, \ldots, x^N) &= \frac{1}{N} \partial_v \partial_x \mathcal{U}\left(s, \frac{1}{N} \sum_{j=1}^N \delta_{x^j}\right)(x^i) + \frac{1}{N^2} \partial^2 \mathcal{U}\left(s, \frac{1}{N} \sum_{j=1}^N \delta_{x^j}\right)(x^i, x^i).
\end{align*}
\]

### 1.5.1 Particle system approximation

Given a McKean–Vlasov SDE on $(\Omega, \mathcal{F}, \mathbb{P})$, a probability space equipped with a filtration $(\mathcal{F}_t)_{t \geq 0}$ and on which we have a $d$–dimensional $\mathcal{F}_t$–Wiener process $(B_t)_{t \geq 0}$ and an independently defined random variable $\xi \in L^2(\mathbb{R}^d)$, once again we consider $(X_j^{\delta_{\xi}}, t \geq 0)$ the solution to (1.4.1). Then, for any $\{\xi_j\}_{j=1}^N$ and $\{B^j\}_{j=1}^N$ independent copies of $\xi$ and $B$ respectively, one can define the following $N$–particle system

\[
\begin{align*}
&\left\{\begin{array}{l}
dY_t^{i,N} = b\left(Y_t^{i,N}, \frac{1}{N} \sum_{j=1}^N \delta_{Y_t^{j,N}}\right) dt + \sigma\left(Y_t^{i,N}, \frac{1}{N} \sum_{j=1}^N \delta_{Y_t^{j,N}}\right) dB_t^i, \quad \forall t \in [0, \infty); \\
y_0^{i,N} = \xi_i,
\end{array}\right.
&i \in \{1, \ldots, N\}. (1.5.2)
\end{align*}
\]

The strong error when approximating (1.4.1) by (1.5.2) is defined for an arbitrary $\phi : \mathbb{R}^d \to \mathbb{R}$ belonging to a certain family of test functions, $N \in \mathbb{N}$ and $T > 0$ as it follows:

\[
\text{strong error} = \left| \mathbb{E}\left[\phi\left(X_T^\xi - \frac{1}{N} \sum_{j=1}^N Y_T^{j,N}\right)\right]\right|.
\]

The weak error, on the other side, is defined as it follows:

\[
\text{weak error} = \left| \mathbb{E}[\phi(X_T^\xi)] - \frac{1}{N} \sum_{j=1}^N \mathbb{E}[^\phi(Y_T^{j,N})]\right|.
\]

The weak and strong errors are alternative approaches for studying the approximation but they in fact complement each other. Indeed, if for example we were to take $\phi(x) = |x|^2$, $\forall x \in \mathbb{R}^d$, while the weak error considers the bias, the strong one considers the variance of the particle approximation. Moreover, the original way of measuring error in such a problem is the strong error and only afterwards, motivated by numerical efficiency, the weak error started to receive some attention. For example, when one applies variance reduction techniques such as antithetic sampling, it is important to study what happens to the bias. In [51], it is shown that simulating $\sqrt{N}$ independent copies of the system with $\sqrt{N}$ particles is less expensive (as soon as the computational cost of the particle system grows more than linearly
in the number of particles) than one simulation of the system with \( N \) particles, although it has bigger bias.

There are available generalisations of these where \( \phi \) is taken to be a function over \( \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \). In this scenario, the weak and strong errors study the approximation of the function of the law of the McKean–Vlasov process by the function evaluated at the empirical measure of the associated \( N \)-particle system.

### 1.5.2 Propagation of chaos

The particle system approximation is justified by the technical result known as propagation of chaos. A detailed, two–parts, survey conducted on the topic of propagation of chaos is [52, 53].

The term propagation of chaos for the particle system (1.5.2) was originally used for denoting the following phenomenon: for any \( \mu_0 \in \mathcal{P}((\mathbb{R}^d)) \) and any choice of deterministic initial states \((Y_{i,0}, N_0)\) for which it is satisfied that \( \mu_{Y_{i,N}} \Rightarrow \mu_0 \) (in the weak sense of probability measures as \( N \to \infty \)), we have the weak limit \( \mu_{Y_{i,N}} \Rightarrow \mu_0 \) in probability in \( \mathcal{P}(\mathbb{R}^d) \), where \( \mu_t = \mathcal{L}(X_t) \), \( \forall \ t \geq 0 \) is the solution to (1.4.1). In other words, the “chaoticity” of the initial distributions (i.e \( \mu_0 \)) “propagates” to later times \( t > 0 \). Relaxation of assumptions over the coefficients and the law of possibly random initial data, in addition to different types of convergence which lead to various explicit rates in \( N \), are two important topics in statistical physics and other areas of applied mathematics.

In [54, Theorem 3.3], the authors present the following qualitative propagation of chaos for Lipschitz coefficients: consider the approximation of (1.4.1) by (1.5.2); then,

\[
(Y_{1,N}, \ldots, Y_{k,N}) \Rightarrow (X_1, \ldots, X_k),
\]

where \( X_1, \ldots, X_k \) are i.i.d. copies of \( X \).

This limit is interpreted as saying that the particles \( \{Y_{i,N}\}_i \) become asymptotically independent, identically distributed as \( N \to \infty \), all of them being copies of the solution to (1.4.1).

There are various techniques used to achieve explicit rates of propagation of chaos, including synchronous coupling, trajectorial propagation of chaos or reflection coupling. An explicit order of strong propagation (particles and limiting copies of the McKean–Vlasov solution belong to the same probability space) was first obtained by Sznitman in [55] (and is stated below) in the additive noise, linear measure dependent drift scenario.

**Theorem 1.5.5.** [55, Theorem 1.4] Given (1.4.1), suppose that \( \xi \in L^2 \), \( \sigma = I_d \) and \( b(\cdot) = \int_{\mathbb{R}^d} \beta(\cdot, x) d\mu(x) \) with \( \beta \) bounded and Lipschitz. Then, for any \( i = 1, \ldots, N \) and \( T > 0 \)

\[
\sup_N \sqrt{N} \mathbb{E} \left[ \sup_{t \leq T} |Y_{i,N}^t - X_i^t| \right] < \infty,
\]

where \( Y_{i,N} \) solve the \( N \)-particle system (1.5.2) and \( X_1, \ldots, X_k \) are i.i.d. copies of \( X \).

Continuing with the strong error, explicit dimension dependent rates were presented in [56]. In terms of dimension–free results, with the help of calculus on measure spaces, [50, Theorem 2.4] obtains for regular enough coefficients and test functions, the uniform rate of \( 1/\sqrt{N} \) for strong propagation of chaos. Another key result is [57, Theorem 2.1], which
presents uniform strong propagation of chaos ((1.4.1) being approximated by (1.5.2)) rates in the framework of constant diffusion and $b = \int \beta(x) \mu(dx)$ with $\beta$ differentiable with Hölder continuous derivative. They suppose moreover that $\mathcal{L}(\xi)$ has a bounded density with respect to the Lebesgue measure and a finite moment of order $2 + \epsilon$ for some $\epsilon > 0$. Then, for any $\rho > 0, T > 0$, they prove the existence of $C < \infty$ such that for all $N \in \mathbb{N}$ and $i \in \{1, \ldots, N\}$,

$$E\left[ \sup_{t \leq T} |Y_{t}^{i,N} - X_{t}^{i}|^{\rho} \right] \leq CN^{-\rho/2},$$

where $Y_{t}^{i,N}$ solve the $N$–particle system (1.5.2) and $X_{t}^{1}, \ldots, X_{t}^{k}$ are i.i.d. copies of $X$. More recently, uniform in time estimates were presented in [58] for dynamics with possibly non–convex confinement and interaction potentials and whose technique relies on a combination of synchronous and reflection couplings with a well adapted (to them) $L^{1}$–weighted Wasserstein metric.

In terms of the weak error, [38, Theorem 2.17] proves the following bias expansion for time–homogeneous coefficients which are $(2k + 1)$–differentiable in space and measure components, bounded diffusion, $\xi \in L^{2k+1}(\mathbb{R}^{d})$ and for measure dependent test functions $\phi$ with the same regularity:

$$E\left[ \phi\left( \frac{1}{N} \sum_{j=1}^{N} Y_{t}^{j,N} \right) - \phi(\mathcal{L}(X_{t}^{i})) \right] = \sum_{j=1}^{k-1} \frac{C_{j}}{N^{j}} + \mathcal{O}\left( \frac{1}{N^{k}} \right),$$

where $C_{j}$ are constants that do not depend on $N$. This results is afterwards used together with Romberg extrapolation to obtain an estimator of $X_{t}^{i}$ with weak error being in the order $\mathcal{O}\left( \frac{1}{N^{k}} \right)$, for each $k \in \mathbb{N}$ (see [38, Section 1.1] for details). Therefore, a higher–order particle system (in terms of the weak error) can be constructed up to a desired order of approximation.

However, uniform estimates are trickier to obtain and have only recently been proved. Optimal rates of uniform in time quantitative strong propagation of chaos for interacting diffusions governed by convex potentials or models on the torus with small interactions, are presented in [59]. Using measure calculus, the authors of [50] proved uniform strong propagation of chaos for regular enough datum, coefficients and test functions (which are assumed Lipschitz). Uniform weak error estimates are presented in [60] for a limited set of examples on the torus by linearizing the Master Equation and relying on calculus on measure spaces.

The other side of this approximating argument is that McKean–Vlasov SDEs are almost impossible to solve explicitly. As a consequence, mathematicians (to be more precise, the pioneer was Boltzmann in the context of statistical physics) simulate these processes through particle systems approximations. In other words, for an implementation point of view the particles systems are more approachable and sampling efficiently has gained traction over the years. Chapter 4 studies subsampling particle systems, which have been proposed as a more efficient technique in approximating the limiting McKean–Vlasov process. In particular, we generalise the rate of convergence in Theorem 1.5.5 to subsampled particle systems (see Theorem C.1.1) and present quantitative weak propagation of chaos in Theorem 4.2.3.

Other relevant references on the topic include [48], [61, Ch.1] and [62].
1.6 Euler scheme

We already talked about implementation hazards and one of the main ones is that although the primary results are obtained in continuous time, one can only put in practice a discretization of the above models. There are many numerical schemes allowing discretization of stochastic process, the main ones being finite elements methods and the Euler–Maruyama scheme. In Chapter 4 we focus on the latter and this section is devoted to it.

Given the stochastic SDE (1.1.2) over the time interval \([0, T] \subset [0, \infty)\), fixed a number of steps \(M \in \mathbb{N}\) and step size \(h > 0\) such that \(T = Mh\), we denote (here and for the rest of this document)

\[ \eta(t) := t_{k-1} = (k-1)h \quad \text{if} \quad t \in [t_{k-1}, t_k), \; k = 1, \ldots, M. \]

The process

\[ \bar{X}_T^{r,s} = x + \int_{\tau}^{T+s} b(\bar{X}_{\eta(r)}^r,x)dr + \int_{\tau}^{T+s} \sigma(\bar{X}_{\eta(r)}^r)dB_r \]  
(1.6.1)

is known as Euler–Maruyama scheme for (1.1.2). Again, if \(\tau = 0\), we use the simplified notation \(\bar{X}_T^r,x = \bar{X}_{\eta(r)}^r, x, \forall r, x\).

A classical result is the one presented in [63] (see also the original [64]), where it is proved that for \(b, \sigma \in C^1_b(\mathbb{R}^d)\) the order of the discretization (strong) error when approximating (1.1.2) by (1.6.1) is \(1/2\), i.e. for every \(T > 0\) there exists \(C > 0\) such that

\[ \mathbb{E} \left[ \sup_{t \leq T} |X_t^x - \bar{X}_T^r,x| \right] \leq Ch^{1/2}. \]

Meanwhile, the order of the weak error is \(\mathcal{O}(h)\) (see [65], which aims at studying the weak rate of convergence under various conditions on the coefficients, including situations where the drift is discontinuous).

Similarly, one considers the Euler scheme for the \(N\)-particle system (1.5.2) as:

\[ \begin{cases} 
 d \tilde{Y}_{i,N}^t = b\left(\tilde{Y}_{i,N}^t, \mu_{\eta(t)} \right)dt + \sigma\left(\tilde{Y}_{i,N}^t, \mu_{\eta(t)} \right)dB_t^i, & i = 1, \ldots, N; \\
 \tilde{Y}_{0,N}^t = \xi_i. 
\end{cases} \]  
(1.6.2)

Strong error discretization rates \(\mathcal{O}(1/\sqrt{N} + \sqrt{h})\) were obtained originally for one-dimensional problems in [66]. Weak error estimates of order \(\mathcal{O}(1/N + h)\) were obtained in [51, Theorem 2.8] for higher dimensions and in the case that the non-linearity in the sense of McKean was given by its moments. Afterwards, these results were generalised in [57, Theorem 2.1] to higher dimensions, discontinuous drifts and constant diffusion. For more general coefficients and measure–dependent test functions, the weak error was proved to be \(\mathcal{O}(1/N + h)\) in [50]. In Proposition C.5.1, we present uniform in time weak error estimates of order \(\mathcal{O}(1/N + h)\) for linear in measure drifts and non–measure–dependent diffusion.

In the same context and after assuming a priori the existence of a function \(\mathcal{R}\) (see Section 4.1.2) representing a bound on the variance of the subsamples of size \(S\) for the law of the state process, we present in Theorem 4.4.2 the following uniform in time weak error order for the Euler scheme subsampled approximating particle system: \(\mathcal{O}(1/N + h + h\mathcal{R}(S)).\)
1.7 Notation

This section is a collection of the notions presented in the present chapter together with a glossary for all the notation needed in the remainder of this dissertation.

1.7.1 Notation

Relevant spaces for the subsequent chapters are the space of $\mathbb{R}^d$–Borel–measurable functions $B(\mathbb{R}^d)$ and, for $m \in \mathbb{N}$, $B_m(\mathbb{R}^d) = \{ \phi \in B(\mathbb{R}^d) \text{ such that there exists } C > 0 \text{ and } |\phi(x)| \leq C(1 + |x|^m), \forall x \in \mathbb{R}^d \}$. Moreover, we consider the space of $p$-continuously differentiable ($p \in \mathbb{N}$) functions on $\mathbb{R}^d$ represented by $C^p(\mathbb{R}^d)$; and for $p_1, p_2 \geq 1$, $I \subseteq \mathbb{R}_+ \equiv [0, \infty)$, the space of $p_1, p_2$-continuously differentiable functions over $I$ and $\mathbb{R}^d$ respectively denoted by $C^{p_1,p_2}(I \times \mathbb{R}^d)$. A $b$ as the subindex, i.e. $C^{p_1,p_2}_b([0, T] \times \mathbb{R}^d)$, will mean that the functions and all the required derivatives are bounded.

For any matrix $M \in \mathbb{R}^{d \times d}$, we denote its transpose by $M^*$ and the component-wise matrix notation is the following: for any $M = (M^{ij})_{i,j=1,...,d} \in \mathbb{R}^{d \times d}$, we denote its rows as $M^i = (M^{ij})_{j=1,...,d} \in \mathbb{R}^d$. Similarly, for any $M = (M^{ij})_{i,j=1,...,d} \in \mathbb{R}^{d \times d}$, we use the following convention: $M^{ij} = (M^{ij})_{i=1,...,d} \in \mathbb{R}^{d \times d}$.

With respect to distances, we denote the $m$–Wasserstein measure by $\mathcal{W}_m$ and in any Euclidean space $|\cdot|$ is used as an arbitrary norm (since they are all equivalent in finite dimensions and our estimates would only change up to a constant). The exceptions are though the trace norm denoted by $\text{tr}()$ and for any $M \in \mathbb{R}^{d \times n}$, the Hilbert–Schmidt norm denoted by $||M||_{HS} = (\sum_{i,j} |M^{ij}|^2)^{1/2}$ and its generalisation, the $L_{p,m}$–norm, $||M||_{p,m} = (\sum_{i} (\sum_{j} |M^{ij}|^p)^{m/p})^{1/m}$. For any tensor $M \in \mathbb{R}^{d \times d \times d}$, we have the equivalent $L_{p,q,m}$–norm,

$$||M||_{p,q,m} = (\sum_{i} (\sum_{j} (\sum_{l} |M^{ijl}|^{q/p})^{m/q})^{1/m}).$$

Moreover, for all $x, y \in \mathbb{R}^d$ we use $xy = \langle x, y \rangle$ to denote the dot product.

Throughout the dissertation, we denote a probability space by $(\Omega, \mathcal{F}, \mathbb{P})$ and, living on it, a $d$-dimensional Wiener process by $(B_t)_{t \geq 0}$. A particularly important role is played by the space of probability measures on $\mathbb{R}^d$ represented by $\mathcal{P}(\mathbb{R}^d)$. We also use $C^{0,0,0}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d))$ to denote the space of functions which are jointly continuous in all its variables (with respect to the relevant metrics in each space) and are $p$–times differentiable in the space variable, uniformly in the measure variables. Additionally, we say $\mathcal{L}(\xi) \in \mathcal{P}(\mathbb{R}^d)$ if for given $W$ a non–negative, growing to infinity, function (i.e Lyapunov function in the sense of Assumption 2.3.3) we have $\int_{\mathbb{R}^d} (1 + W(x)) \mathcal{L}(\xi) \text{d}x < \infty$. On the other hand, for any function $\phi \in B(\mathbb{R}^d)$, we say $\phi \in L^p(\mathbb{Q})$ when $\int_{\mathbb{R}^d} |\phi(x)|^p \mathbb{Q} \text{d}x < \infty$. The measure is only omitted when we refer to the Lebesgue measure. Similarly, an $\mathbb{R}^d$–valued random variable $\xi$ is said to belong to $L^p(\mathbb{R}^d)$ if $\mathbb{E}|\xi|^p < \infty$.

Important notions are $1_A$ representing the index function of the set $A \subseteq \mathbb{R}^d$ and the Dirac measure for a given $x \in \mathbb{R}^d$ denoted by $\delta_x$. They are defined for any $x \in \mathbb{R}^d$ and $A \subseteq \mathbb{R}^d$ as

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

With respect to distances in measure spaces, we denote the $m$–Wasserstein measure by $\mathcal{W}_m$ and the Total Variation norm by $||\cdot||_{TV}$ (see [11, pp. 436, 244]). Additionally, recall [3] that for a given function $W$ on $[0, \infty)$, the Weighted Total Variation (WTV) norm, $||\cdot||_W$, and WTV

$$30$$
distance, \( d_W \), are given for any two measures \( q, q' \in \mathcal{P}^W(\mathbb{R}^d) \) by
\[
\|q\|_W := \int_{\mathbb{R}^d} (1 + W(x)) d\mu(x)
\]
\[
d_W(q, q') := \int_{\mathbb{R}^d} (1 + W(x))|q - q'|(x) = 2\|((1 + W(\cdot))(q - q'))\|_{TV} = \|q - q'\|_W.
\]

Next, let us mention that, if they exist, we denote the \( n^{th} \)-derivatives in the first space component of the function \( \psi : (0, \infty) \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \) by \( \partial_{x_i}^n \psi(t, x, \xi) = \partial_{x_i}^n \psi(t, x, \xi) \), where \( x = (x^i)_{i=1,...,d} \in \mathbb{R}^d \) and \( \alpha = (\alpha_i)_{i=1,...,d} \in \mathbb{N}^d \) any multi-index such that \( |\alpha| = n \). This is an abuse of notation justified by the fact that we only care about the derivative’s order. Similarly, \( \partial_{\xi_j}^n \psi(t, x, \xi) \) is used for the derivative in second space component and \( \partial_t \psi(t, x, \xi) \) for the derivative in the time component. Now, if the function being differentiated has only one space component, e.g. \( \psi : (0, \infty) \times \mathbb{R}^d \to \mathbb{R} \), then we adopt to the following notation: \( \partial_{x_i}^n \psi(t, x) \), where for \( n = 1 \) we obtain the gradient \( (\partial_{x_i} \psi(t, x))_{i=1,...,d} \) and for \( n = 2 \) the Hessian \( (\partial_{x_i, x_j}^2 \psi(t, x))_{i=1,...,d} \).

Finally, let us mention that throughout the dissertation \( C > 0 \) is a constant changing value from line to line, which might be dependent on the dimension, the coefficients and the remaining parameters in the set of assumptions but crucially independent of time (noted with \( s, t \) or \( r \) and space (noted by \( x, x', x'', y, z \) or even \( \xi \)).

### 1.8 Overview

This section is devoted to summarizing what is studied in this thesis and is developed in the subsequent three chapters. This is done in the reversed order with respect to the numbering of the chapters, in order to present a motivational overview of the developed theory. We also explain how they are related to each other and to the literature.

In Chapter 4, consisting of the first project, we present a detailed study of the subsampling technique. Subsampling is used broadly in the machine learning field but it lacks a fully rigorous mathematical analysis. We present a stochastic framework in order to do so, analyse it using different examples and obtain uniform weak error estimates induced by approximating McKean-Vlasov SDEs with subsampled particle systems. For non–subsampled approximating particle systems, there are already results in the literature regarding the weak error estimates (see [50, 57, 67]). We generalise these to McKean–Vlasov dynamics and introduce an extra dependence on an external measure in order to accommodate applications to 1–layer Neural Networks (Theorem 4.2.1). Furthermore, we present weak error estimates when one considers Euler–Maruyama time discretization of the subsampled particle system (Theorem 4.4.2). A key result for attaining uniformity in time using our result is the decaying derivative estimates for the backward Kolmogorov function associated to the underlying McKean–Vlasov process. In other words, first we must prove time–decaying derivative estimates for the solution to the associated backward Kolmogorov equation. In the same chapter, we obtain these under monotonicity assumptions by introducing the so called derivative processes (Proposition 4.3.10) and relying on the regularity of the test function.

This application led us to a more detailed study of the solution to the backward Kolmogorov equation and we noticed the absence of results regarding the time behaviour of its derivatives. The available literature, to the best of the author’s knowledge, achieve time decay
of such derivatives only in the time–homogeneous and smooth coefficients; or finite horizon scenarios ([25–27, 68] some which have been collected in Section 1.3.2). In Chapter 2, we present a technical result on the time behaviour of as many derivatives as required (Theorem 2.2.4) under two essential assumptions but without requiring differentiability of the test function. One of the hypotheses is that the transition density of the non–autonomous stochastic measure decays or “sticks” to a static measure (Assumption 2.2.2). This determine later on whether the derivative estimates decay or are uniform in time. The second hypothesis consists of the existence of enough derivatives for the transition density (not necessarily infinitely many) and a certain integrability property satisfied by these (Assumption 2.2.3). In addition, we present a possible use of this result to an example satisfying a better studied set of conditions, applied directly on the coefficients.

Finally, Chapter 3 generalises the previous results in order to obtain gradient estimates for backward Kolmogorov functions associated to McKean–Vlasov processes. Sufficient conditions are provided for its decay in time in Corollary 3.4.1. However, since the underlying process is McKean–Vlasov, the gradient is is decomposed in two. One component can be dealt with by providing sufficient conditions for Theorem 2.2.4 (applied to non-autonomous SDE) to hold. And in fact, we obtain not only first but also second order derivatives decay for this part of the derivative. And another component which is inherently McKean and which we approach using a modified Bismut–Elworthy–Li formula (Theorem 3.2.7) and moments estimates based on monotonicity (see Assumption 3.2.3). These estimates for derivatives of the backward Kolmogorov function associated to a McKean–Vlasov process, and second order derivatives ones as well, have only been achieved recently in the context of measure space calculus and for particular types of dynamics on the torus (see [60]).

To conclude, in this thesis we have shown various sets of assumptions under which we conclude time–decay of the derivatives of the solution to the backward Kolmogorov equation associated to both: non–autonomous SDEs and McKean–Vlasov SDEs. These allow us to derive uniform weak error estimates for subsampled approximating particle systems.
Chapter 2

Decaying derivative estimates for functions of solutions to non-autonomous SDEs

We produce uniform and decaying in time bounds for derivatives of the solution to the backwards Kolmogorov equation associated to a stochastic processes governed by a time dependent dynamics. These hold under assumptions over the integrability properties in finite time of the derivatives of the transition density associated to the process, together with the assumption that we have some control over the distance between the transition density and a static measure. We moreover provide examples which satisfy such a set of assumptions.

2.1 Introduction

In this chapter we consider the following real–valued, $d$-dimensional stochastic process $(X_t)_{t \geq 0}$ satisfying a non–autonomous SDE. Indeed, given time–dependent coefficients $\beta : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}^d$ and $\Sigma : [0, \infty) \times \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ and initial datum $x \in \mathbb{R}^d$, our object of study is the stochastic process $(X^x_t)_{t \geq 0}$ assumed to be the unique (in the sense of probability law) weak solution of the following SDE with Brownian motion process $(B_s)_{s \geq 0}$:

$$dX^x_t = \beta(s, X^x_s)\, ds + \Sigma(s, X^x_s)\, dB_s, \quad \forall s \in (0, \infty); \quad X^x_0 = x. \quad (2.1.1)$$

Let $\phi : \mathbb{R}^d \to \mathbb{R}$ be a measurable function. Then, due to uniqueness in law of solution to (2.1.1), $(0, \infty) \times \mathbb{R}^d \ni (s, x) \mapsto V(s, x) := \mathbb{E}[\phi(X^x_s)]$ is a well defined function under different conditions over $\phi, \beta, \Sigma$ (see [5, Corrollary 2.4] for coefficients satisfying the conditions in Theorem 1.1.14 and $\phi \in C^\infty_b(\mathbb{R}^d)$; or [5, Remark 5.9] for coefficients satisfying the global analogue of the conditions in Theorem 1.1.14 and $\phi$ with polynomial growth). Moreover it satisfies, under enough assumptions for the coefficients, a certain PDE known as the backward Kolmogorov equation (see e.g. [1, Theorem 7.6] or [5]).

For only space dependent, smooth coefficients $\beta, \Sigma$ with bounded derivatives of any order, in addition of $\Sigma$ being bounded itself; and smooth function $\phi$, the authors of [27] were motivated by uniform weak error for Euler scheme estimates to obtain exponential decay for derivatives of $V$ of any order using Sobolev embeddings. For strictly space dependent,
bounded and Hölder continuous coefficients with derivatives up to certain order, we have the results of [28] which are heavily based on previously obtained rate of convergence to invariant measures in [69]. More recent work was done by Menozi, Pesce and Zhang [70] for uniform bounds for \( \partial_x V(s,x) \) and \( \partial^2_x V(s,x) \), where the coefficients are assumed Hölder continuous in space and the drift has linear growth.

The novelty of our results is that we translate finite time results to infinite time. More precisely, we obtain explicit estimates in the time interval \([0, \infty)\) of the space derivatives of \( V \) up to the order imposed by the regularity and integrability of the transition density of the SDE (2.1.1) in a fixed finite time interval. The order of decay is given by a weaker condition than decay to invariant measure (see Assumption 2.2.2).

An application is the use of such derivative estimates to obtain explicit weak error rates when approximating a process \((X_{t+s}^x)_{s \geq 0}\) by an Euler scheme. Namely, the definition of \( V \) and the initial data of the PDE (2.1.1), allows us to recast the expression of the weak error into another one to which we can apply Itô’s formula. Additionally, the use of the backward Kolmogorov equation simplifies the expression result of this computation and lets us split the weak error into more approachable terms. Together with decaying in time bounds on the space derivatives of \( V \), these allow obtaining uniform such weak error orders. In the case that \((X_{t+s}^x)_{s \geq 0}\) follows a McKean–Vlasov dynamics (for more information see Chapter 3), the decaying in time estimates on the space derivatives of \( V \) allow obtaining explicit, uniform in time weak error order for approximating particle systems (see [38, 57]).

The chapter is organised as follows: in Section 2.2 we formulate conditions under which we prove that the desired derivative estimates hold. They are presented in the first main result of the chapter: Theorem 2.2.4. Its proof is inspired by [28] in the sense that we also use Chapman–Kolmogorov identity which allows us to move the derivatives of \( V \) onto the derivatives of the transition density of \( X \) in finite time intervals. In Section 3 we then show two examples satisfying this set of assumptions in the non-autonomous SDE setup and we state the other main result, which brings down to earth the results in Section 2.2 and is presented in Theorem 2.3.9.

### 2.2 Main result: Derivative with respect to the initial condition

For \( x \in \mathbb{R}^d \) and \( \tau \geq 0 \), we consider the following non-autonomous SDE

\[
X_{t+s}^{x,\tau} = x + \int_T^{T+s} \beta(t, X_t^{x,\tau}) \, dt + \int_T^{T+s} \Sigma(t, X_t^{x,\tau}) \, dB_t, \quad s \geq 0.
\]

(2.2.1)

In order to avoid cumbersome notation, we will be using the following convention \( X_{0+s}^{x,0} = X_s^x \) for all \((s,x) \in [0, \infty) \times \mathbb{R}^d\).

To formulate the assumptions let us fix \( \mathcal{F} \subseteq B(\mathbb{R}^d) \) and \( N \in \mathbb{N} \).

**Assumption 2.2.1** (Conditions for the density of the transition probability). For all \( x \in \mathbb{R}^d \), \( \tau \geq 0 \), the equation (2.2.1) has a solution \((\Omega, \mathbb{P}, (\mathcal{F}_s)_{s \geq \tau}, (X_{s}^{x,\tau})_{s \geq \tau}, (X_{s}^{x,\tau})_{s \geq \tau})\) unique in the sense of probability.

Moreover, we suppose that this process admits a density: \((s,x') \mapsto p_{t+s}^x(x,x')\).
Finally, either $\mathcal{S} \subseteq C(\mathbb{R}^d)$ or for any $0 \leq n \leq N$ there exists $\delta > 0$ such that
\[
\sup_{\phi \in \mathcal{S}} \int_{\mathbb{R}^d} |\phi(x'')| \partial^n_x p^n_0(x, x'')^{1+\delta} \, dx'' < \infty, \quad \forall (s, x) \in (0, \infty) \times \mathbb{R}^d.
\]

**Assumption 2.2.2 (Sticking to a measure).** There exist $q \in \mathcal{P}(\mathbb{R}^d)$, $g : \mathbb{R}^d \to \mathbb{R}_+$ and $G : \mathbb{R}_+ \to \mathbb{R}_+$ such that for all $(s, x) \in (0, \infty) \times \mathbb{R}^d$:
\[
\sup_{\phi \in \mathcal{S}} \left| \int_{\mathbb{R}^d} \phi(x') \left( p_{1+s}^1(x, dx') - q(dx') \right) \right| \leq g(x) G(s).
\]

Assumption 2.2.2 deserves a few comments. First, in most scenarios we would like $\lim_{s \to \infty} G(s) = 0$, i.e. the transition density is not only sticking to a static measure but decaying to one. Next, by considering a tailor–made family of test functions, one can make use of familiar metrics in order to verify Assumption 2.2.2. Indeed, suppose that $\mathcal{S}$ is formed only by 1–Lipschitz functions then it is enough for $q$ to be a limit in the 1–Wasserstein distance of the law of the process satisfying (2.2.1). Yet another possibility is to consider $\mathcal{S}$ as the family of locally Lipschitz functions and have convergence of the solution to (2.2.1) to $q$ in the corresponding Wasserstein distance (see Lemma 2.3.9).

A key remark is that $q$ is just a limit measure (if $G$ decreases to 0), and with Assumption 2.2.2 we are not covertly asking for the existence of a unique invariant measure in any of the mentioned distances. Moreover, if $G$ in not a decreasing function but merely a bounded one, this assumption requires the law of the the solution to (2.2.1) to stay “close” to the fixed measure $q$, without having to converge to it. Consequently, the space derivatives of the functions $V$ can only be concluded to be uniform (but not decaying) in time when applying Theorem 2.2.4.

**Assumption 2.2.3 (Smoothness and integrability of derivatives of the density w.r.t. the starting point).** For any $0 \leq n \leq N$, we assume $(s, x', x) \to \partial^n_x p^n_0(x, x')$ exist and are continuous in $(x, x')$. Moreover, there exists $h : \mathbb{R}^d \to \mathbb{R}_+$ such that for any $1 \leq n \leq N$ and $g$ satisfying Assumption 2.2.2, the following is satisfied:
\[
\int_{\mathbb{R}^d} g(x'') |\partial^n_x p^n_0(x, x'')| \, dx'' \leq h(x), \quad \forall x \in \mathbb{R}^d.
\]

**Theorem 2.2.4.** Let $\phi \in \mathcal{S}$, $(X^x_s)_{s \geq 0}$ be the unique (in law) solution of (2.2.1) and
\[
V(s, x) := \int_{\mathbb{R}^d} \phi(x') p^n_0(x, x') \, dx' = \mathbb{E}[\phi(X^x_s)].
\]

If Assumptions 2.2.1, 2.2.2 and 2.2.3 hold, then for all $1 \leq n \leq N$, we have
\[
|\partial^n_x V(s, x)| \leq h(x) G(s), \quad \text{for all} \quad (s, x) \in (1, \infty) \times \mathbb{R}^d.
\]

Note that, if $\lim_{s \to \infty} G(s) = 0$ we conclude that the derivatives w.r.t. $x$ of $V(s, x)$ decay to zero.
Proof. Let us first show that due to Assumption 2.2.1, for all \( 1 \leq n \leq N \) and for all \( x \in \mathbb{R}^d \)

\[
\partial_x^n V(s, x) = \partial_x^n \int_{\mathbb{R}^d} \phi(x') p_0^0(x, x')dx' = \int_{\mathbb{R}^d} \phi(x'') \partial_x^n p_s^0(x, x'')dx''.
\] (2.2.2)

Indeed, if \( \phi \) is continuous, we can differentiate under the integral sign with Leibniz formula (see [18, Theorem 12.14]). Otherwise we argue as follows. Recall that for all \( 1 \leq n \leq N \) and for all \( x \in \mathbb{R}^d \) we assumed the existence of a \( \delta > 0 \) such that \( \int_{\mathbb{R}^d} |\phi(x'')|^{1+\delta} |\partial_x^n p_s^0(x, x'')|^{1+\delta} dx'' < \infty \). As a consequence, for any order \( 1 \leq n \leq N \) and \( h > 0 \), \((s, x^n) \in (0, \infty) \times \mathbb{R}^d \) and any element in an orthonormal basis in \( \mathbb{R}^d \) represented as \( \{e_i\}_{1,...,d} \),

\[
\sup_{h \geq 0} \int_{\mathbb{R}^d} |\phi(x'') \frac{1}{h} (\partial_x^{n-1} p_s^0(x + he_i, x'') - \partial_x^{n-1} p_s^0(x, x''))|^{1+\delta} dx'' \\
\leq \int_{\mathbb{R}^d} |\phi(x'')|^{1+\delta} \sup_{h \geq 0} \frac{1}{h} (\partial_x^{n-1} p_s^0(x + he_i, x'') - \partial_x^{n-1} p_s^0(x, x''))|^{1+\delta} dx'' \\
\leq \int_{\mathbb{R}^d} |\phi(x'')|^{1+\delta} |\partial_x^n p_s^0(x, x'')|^{1+\delta} dx'' < \infty.
\]

Meaning that by De La Vallée Poussin Theorem (see [22, Theorem 2.4.4]), for any order \( 1 \leq n \leq N \) and for all \( x \in \mathbb{R}^d \), the increments

\[
\phi(x'') \frac{1}{h} (\partial_x^{n-1} p_s^0(x + he_i, x'') - \partial_x^{n-1} p_s^0(x, x'')) , \quad i = 1,...,d;
\]

are uniformly (in \( h \)) integrable in \( x'' \) over \( \mathbb{R}^d \). This means that we can apply Vitali’s Convergence Theorem (see [17, Chapter 4]) and obtain (2.2.2) by induction from:

\[
\lim_{h \to 0} \int_{\mathbb{R}^d} \phi(x'') \frac{\partial_x^{n-1} p_s^0(x + he_i, x'') - \partial_x^{n-1} p_s^0(x, x'')}{h} dx'' \\
= \int_{\mathbb{R}^d} \phi(x'') \lim_{h \to 0} \frac{\partial_x^{n-1} p_s^0(x + he_i, x'') - \partial_x^{n-1} p_s^0(x, x'')}{h} dx''.
\]

Let \( s \geq 1 \) and recall the non-autonomous Chapman–Kolmogorov identity

\[
p_{1+s}^0(x, x'', t) = \int_{\mathbb{R}^d} p_{1+t}^0(x', x') p_{1+s}^0(x', x'') dx',
\] (2.2.3)

whose proof we include in Appendix A.1. Let us now apply it with \( \tau = 0 \) in (2.2.2) for any \( 1 \leq n \leq N \) and \( x \in \mathbb{R}^d \). After taking the derivatives inside the second integral with Leibniz formula (this is allowed given the continuity of \((x, x') \mapsto \partial_x^n p_0^0(x, x')\) stated in Assumption 2.2.2 for \( 0 \leq n \leq N \), see [18, Theorem 12.14]), we obtain:

\[
|\partial_x^n V(s, x)| = \left| \int_{\mathbb{R}^d} \phi(x'') \partial_x^n p_s^0(x, x'') dx'' \right| = \left| \int_{\mathbb{R}^d} \phi(x'') \partial_x^n ( \int_{\mathbb{R}^d} p_0^0(x, x') p_{1+s}^0(x', x'') dx' ) dx'' \right| \\
= \left| \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \phi(x'') \partial_x^n p_0^0(x, x') p_{1+s}^0(x', x'') dx' dx'' \right|.
\]

Now notice that first by Fubini’s Theorem, afterwards by Leibniz formula due to continuity of \( \partial_x^n p_0^0(x, x') \) for \( 0 \leq n \leq N \) and finally by the fact that \( q \) is independent of the initial data, we
conclude:
\[ \int_{\mathbb{R}^d} \phi(x') \partial_x^n p_0^1(x, x') q(dx'') \leq \int_{\mathbb{R}^d} \phi(x'') \partial_x^n \left( \int_{\mathbb{R}^d} p_0^1(x', dx') q(dx'') \right) \]
\[ = \int_{\mathbb{R}^d} \phi(x'') \partial_x^n \left( \int_{\mathbb{R}^d} p_0^1(x', dx') \right) q(dx'') \]
\[ = \int_{\mathbb{R}^d} \phi(x'') \partial_x^n (1) q(dx'') \quad (2.2.4) \]
\[ = 0. \]

This allows us to continue the above chain of equalities as
\[ |\partial_x^n V(s, x)| = |\int_{\mathbb{R}^d} \phi(x'') \partial_x^n p_0^1(x, x') \left( p_1^{t+s}(x', dx'') - q(dx'') \right) dx'| \]
\[ \leq \int_{\mathbb{R}^d} \partial_x^n p_0^1(x, x') \left| \int_{\mathbb{R}^d} \phi(x'') \left( p_1^{t+s}(x', dx'') - q(dx'') \right) dx' \right. \]

Since \( \phi \in \mathcal{S} \) and due to Assumption 2.2.2 together with Assumption 2.2.3, we conclude that
\[ |\partial_x^n V(s, x)| \leq \int_{\mathbb{R}^d} |\partial_x^n p_0^1(x, x')| g(x') G(s) dx' \]
\[ \leq h(x) G(s). \]

This completes the proof. \( \square \)

To finish this section, let us enumerate a few well studied possibilities of obtaining Assumptions 2.2.1, 2.2.2 and 2.2.3, which guarantee the estimates in Theorem 2.2.4.

First, one can extract derivative bounds of the transition density of \( (X_{t+\tau})_{s \in [\tau, T]} \) in finite time from the PDEs literature. Many results are available in this direction when we allow the time interval to be of a fixed length \( T - \tau > 0 \): Friedman in [25] and Eidelman in [26] are two main references. Their restrictions come from the smoothness required for the coefficients: bounded and uniformly Hölder continuous for Friedman and bounded diffusion and linearly growing drift for Eidelman. An extension of the later was obtained recently by Menozzi, Pesce and Zhang in [70]. In this later result the diffusion is assumed to be Hölder continuous in space and the drift to have linear growth. For a general study of transition densities, see [8].

Second, one can classify the methods for obtaining decay to the invariant measure, if what we are after are decaying derivatives bounds, in three categories. For each, we include a single reference and the reader is remitted to [30] for a more detailed list of references for: (i) the approach based on Harris theorem or the Meyn–Tweedie approach with Lyapunov functions [31]; (ii) the approach based on entropy estimates and Poincaré and Sobolev inequalities [32]; and (iii) the probabilistic approach based on coupling [33].

### 2.3 Application to non–autonomous SDEs

This section is dedicated to finding a tractable set of assumptions which imply in turn Assumptions 2.2.2 and 2.2.3. Provided the road map in the previous section, we show two alternatives for obtaining each of the mentioned assumptions, any of which when combined
We assume they are uniformly elliptic over $[0,1] \times \mathbb{R}^d$ and one of the following regularity and growth conditions holds for the coefficients:

- The coefficients $\beta, \Sigma \in C^{1,2}_b([0,1] \times \mathbb{R}^d)$, $\beta, \Sigma \in C^{1,2}([0,\infty) \times \mathbb{R}^d)$ and $\Sigma$ is bounded on the whole $[0,\infty) \times \mathbb{R}^d$. We denote by $M > 0$ the bound on the diffusion.

- The diffusion $\Sigma \in C_b^{0,2}([0,\infty) \times \mathbb{R}^d)$. Moreover, $\Sigma$ is continuous in $t$ uniformly in $x$. For the drift we assume $\beta \in C^{0,2}([0,\infty) \times \mathbb{R}^d)$.

Additionally we assume that there exist $M_0 \geq 0$ and $0 < \lambda < 1$ such that for all $t \in [0,1]$ and $x, y \in \mathbb{R}^d$, \[
|\beta(t,x) - \beta(t,y)| + |\Sigma(t,x) - \Sigma(t,y)| \leq M_0|x - y|^\lambda;
\]
and there exist $M, \epsilon > 0$ such that for all $(t,x) \in [0,1] \times \mathbb{R}^d$, \[
|\partial_x \beta(t,x)| \leq M(1 + |x|)^{1-\epsilon}, \quad |\partial_x^2 \beta(t,x)| \leq M(1 + |x|)^{3-\epsilon};
\]
and
\[
|\Sigma(t,x)| \leq M, \quad |\partial_x \Sigma(t,x)| \leq M(1 + |x|)^{1-\epsilon}, \quad |\partial_x^2 \Sigma(t,x)| \leq M(1 + |x|)^{2-\epsilon}.
\]

Notice that $\lambda$–Hölder continuity and continuity in the time variable $(t)$ imply sub–linear growth: there exists $M > 0$ such that $|\beta(t,x)| \leq M(1 + |x| \lambda^2)$. In particular, $|\beta(t,x)| \leq M(1 + |x|)$.

### Assumption 2.3.2 (Uniform ellipticity)

Let us consider the differential operators $L, L^*$ defined by
\[
C^2(\mathbb{R}^d) \ni u \rightarrow Lu = \frac{1}{2} \text{tr}(\Sigma \Sigma^\ast \partial_x^2 u) + \beta \partial_x u \quad \text{and} \quad C^2(\mathbb{R}^d) \ni u \rightarrow L^* u = \frac{1}{2} \text{tr}(\partial_x^2 (\Sigma \Sigma^\ast u)) + \partial_x (\beta u).
\]
We assume they are uniformly elliptic over $[0,\infty) \times \mathbb{R}^d$, i.e. there exists $\kappa > 0$ such that for all $(t, x, \xi) \in [0,1] \times \mathbb{R}^d \times \mathbb{R}^d$, we have $\xi \ast \Sigma \Sigma^\ast(t,x) \xi \geq \kappa |\xi|^2$.

### Assumption 2.3.3 (Lyapunov function)

There exists a function $W(x) : \mathbb{R}^d \rightarrow [0,\infty)$ such that $W \in B_p(\mathbb{R}^d)$ for some $p \in \mathbb{N}$, $\lim_{|x| \rightarrow \infty} W(x) = \infty$ and $W \in C^2(\mathbb{R}^d)$. Moreover, there exists $M_W > 0$ such that $|\partial_x W(x)| \leq M_W(1 + W(x))$, $\forall x \in \mathbb{R}^d$ and there exists $M_1 > 0$ such that for all $t \in (1,\infty)$, $x, y \in \mathbb{R}^d$,
\[
(L(t,x) - L(t,y)) W(x - y) \leq -M_1 W(x - y).
\]
Notice that a natural choice for such a Lyapunov function is $W(x) = |x|^p$ for some $p \geq 1$.

Finally, for a fixed value $m \in \mathbb{N}$, we define our family of test functions
\[
\mathcal{S}_m := \{ \phi : \mathbb{R}^d \rightarrow \mathbb{R} | \phi \in B(\mathbb{R}^d) \text{ and } |\phi(x) - \phi(y)|^m \leq W(x - y), \forall x, y \in \mathbb{R}^d \}.
\]
Next we present a few, even more explicit, examples where Assumptions 2.3.2, 2.3.1 and 2.3.3 are satisfied simultaneously.

**Example 2.3.4.** The classical example when \( W(x) = |x|^2, \forall x \in \mathbb{R}^d \) and the drift is allowed to have linear growth is: \( 0 < \Sigma \in \mathbb{R}^d \) and \( [0, \infty) \times \mathbb{R}^d \ni (t, x) \rightarrow \beta(t, x) = -M_1 x \) for some \( M_1 > 0 \). Then naturally, Assumptions 2.3.1(b) and (2.3.3) are satisfied on the whole \([0, \infty) \times \mathbb{R}^d\).

**Example 2.3.5.** In \( d = 1 \), again constant diffusion \( \Sigma \in (0, \infty), (0, \infty) \times \mathbb{R} \ni (t, x) \rightarrow \beta(t, x) = \sin(x) - (M_1 + 1)x \) for some \( M_1 > 0 \) (any bounded, with negatively bounded derivative function would work). This example satisfies Assumption 2.3.1(a) and moreover, by the Mean Value Theorem, Assumption 2.3.3 holds with \( W(x) = |x|^2, \forall x \in \mathbb{R}^d \). More explicitly, for any \( t \geq 0 \) and \( x, y \in \mathbb{R} \), there exists \( \xi \in \mathbb{R} \) such that

\[
(L(t, x) - L(t, y))W(x - y) = 2(\beta(t, x) - \beta(t, y))(x - y) = 2\partial_x \beta(t, \xi)(x - y)^2 \leq -2M_1 W(x - y).
\]

**Example 2.3.6.** In \( d = 1 \), \( \Sigma \in (0, \infty) \) and \( (0, \infty) \times \mathbb{R} \ni (t, x) \rightarrow \beta(t, x) = -xe^{-tx + 100(t-1)x^2 - 10(1.1-t)x^2}. \)

These coefficients are obviously continuous. Moreover, notice that on one hand, for \( t \in [0, 1) \) we have that \( \beta(t, x) \) is bounded and therefore this example satisfies Assumption 2.3.1(a). And on the other hand,

\[
\sup_{x \in \mathbb{R}, t \geq 1} \partial_x \beta(t, x) = \sup_{x \in \mathbb{R}} -1 - xt + 200x^2(t-1) - 20x^2(1.1-t)e^{-tx + 100(t-1)x^2 - 10(1.1-t)x^2} \leq 0.
\]

And since the diffusion is a constant, Assumption 2.3.3 is satisfied for \( t \geq 1 \) and \( W(x) = x^2, \forall x \in \mathbb{R} \), only that for \( M_1 = 0 \). Hence this example does not fall exactly under our assumptions. Instead, if one allows for \( M_1 \) in Assumption 2.3.3 to be time dependent such that \( M_1(t) \geq 0 \) for all \( t > 0 \) such that \( M_1(t) = 0 \) if and only if \( t = 0 \), all our computations are valid. In order to avoid confusing the message, we proceed for the remainder of the dissertation with constant \( M_1 > 0 \).

The idea one must take away from this example is that once the time dependence was the limitation to applying results in the literature, but now the dependence on the time variable is what drives the shift in between the bounded and the “decaying as \(-x\)” behaviour.

### 2.3.1 Finite time estimates on derivatives of transition densities

The forward Kolmogorov equation associated to the process solving the SDE (2.2.1) with initial time \( \tau \geq 0 \) and initial data \( X_\tau^T = z \in \mathbb{R}^d \) is the following PDE

\[
\partial_t p_t^T(z, x) - L^*(t, x)p_t^T(z, x) = 0, \quad (t, x) \in [\tau, \infty) \times \mathbb{R}^d.
\] (2.3.2)

There are many relevant results on well posedness and solution regularity. First, existence of solution follows from the existence of solution to (2.2.1) under Assumptions 2.3.1 and 2.3.3. Uniqueness can be guaranteed under one of the conditions in Assumption 2.3.1 (see [2, Theorem 7.4] or [25]). Next we are going to enumerate a few results regarding the stability, regularity and explicit estimates of such solutions collected from the literature on parabolic PDEs and adapt them to the shape of the assumptions in Section 2.2.

**Lemma 2.3.7.** Let Assumptions 2.3.2 and 2.3.1(a) hold. Assume moreover that \( W \) satisfies that

\[
\sup_{c > 0} \int_{\mathbb{R}^d} \mathbb{E}|W(x)|e^{-c|x|^2} \, dx < \infty.
\]

For any \( g : \mathbb{R}^d \rightarrow \mathbb{R}_+ \) such that there exists \( C > 0, m \in \mathbb{N} \) for which
g(x) ≤ C(1 + W(x))^{1/m}, and φ ∈ ℋ_m (defined in (2.3.1)), Assumption 2.2.3 is satisfied with \( h(x) = C e^{c|x|^2} \), for all \( x \in \mathbb{R}^d \). Moreover, if \( W \in B_p(\mathbb{R}^d) \) then Assumption 2.2.1 holds for some \( δ > 0 \) and there exist \( C, c > 0 \) such that Assumption 2.2.3 is satisfied with

\[
h(x) = \begin{cases} 
C e^{c|x|^2}, & \text{if } p < m \\
C(1 + |x|^{p/m}), & \text{if } p \geq m;
\end{cases}
\]

for all \( x \in \mathbb{R}^d \).

**Proof.** From Assumption 2.3.2 and Assumption 2.3.1(a), all the conditions in [25, Theorem 9.4.2, Remark below display (9.418)] are satisfied and hence for \( 0 ≤ n ≤ 2 \), there exist \( C, c > 0 \) such that:

\[
|\partial_x^n p_1^0(x, x'')| ≤ C e^{-c|x-x''|^2}, \quad \forall x, x'' \in \mathbb{R}^d.
\]

Moreover, for all \( t > 1 \) and \( x, x'' \in \mathbb{R}^d \), by Girsanov’s theorem,

\[
|p_1^1(x, x'')| ≤ \frac{C}{(t-1)^{d/2}} e^{-c\left(\frac{|x-x''|^2}{t-1}\right)}.
\]

Then we have directly Assumption 2.2.3 with \( \mathbb{R}^d \ni x \mapsto h(x) = C e^{c|x|^2} \) (with some other \( C, c > 0 \)) since by Young’s inequality we have \(|x||x''| ≤ |x|^2 + \frac{1}{4}|x''|^2\) and therefore

\[
\int_{\mathbb{R}^d} g(x'')|\partial_x^n p_0^0(x, x'')|dx'' ≤ C \int_{\mathbb{R}^d} (1 + W(x''))^{1/m} e^{-c|x-x''|^2} dx''
\]

\[
≤ C \int_{\mathbb{R}^d} (1 + W(x''))^{1/m} e^{-c|x|^2 - c|x''|^2 + 2c|x||x''|} dx''
\]

\[
≤ C \int_{\mathbb{R}^d} (1 + W(x''))^{1/m} e^{-c|x|^2 - c|2|x''|^2} dx''
\]

\[
≤ C e^{c|x|^2}.
\]

In particular, if \( W \in B_p(\mathbb{R}^d) \),

\[
\int_{\mathbb{R}^d} (1 + W(x''))^{1/m} e^{-c|x''|^2} dx'' ≤ C \int_{\mathbb{R}^d} (1 + |x''|^{p/m}) e^{-c|x''|^2} dx'' ≤ C.
\]

Notice however that if \( p > m \) then by a change of variables and noticing that \(|x + x''|^{p/m} ≤ C(|x|^{p/m} + |x''|^{p/m})\), we can conclude an improved estimate:

\[
\int_{\mathbb{R}^d} g(x'')|\partial_x^n p_1^0(x, x'')|dx'' ≤ \int_{\mathbb{R}^d} C(1 + W(x''))^{1/m} e^{-c|x-x''|^2} dx''
\]

\[
≤ C \int_{\mathbb{R}^d} (1 + |x'|^{p/m} + |x''|^{p/m}) e^{-c|x'|^2} dx' \leq C(1 + |x|^{p/m}).
\]

We focus now on Assumption 2.2.1. Let \( δ > 0 \) be arbitrary. By Chapman–Kolmogorov identity, Leibniz formula and Jensen's inequality:

\[
\int_{\mathbb{R}^d} |\phi(x'')|^{1+δ} dx'' ≤ \int_{\mathbb{R}^d} |\phi(x'')|^{1+δ} \left|\int_{\mathbb{R}^d} p_1^1(x', x'') \phi(x') dx'\right|^{1+δ} dx''
\]

\[
≤ \int_{\mathbb{R}^d} |\phi(x'')|^{1+δ} \left|\int_{\mathbb{R}^d} p_1^0(x', x'') \partial_x^n p_1^0(x, x') dx'\right|^{1+δ} dx''
\]
Moreover, there exists a constant $C_r > 0$ (again changing value from line to line but also dependent on $r$) such that for $n$, $p \leq C_r$ as a direct consequence of the results in [25]. Similarly, there exists $C > 0$ such that for $n = 1, 2$, we have $\sup_{x, x'} p^n(x, x') \leq C$. Therefore, if $\phi \in \mathcal{S}_m$ and $W \in B_p(\mathbb{R}^d)$, we conclude by (2.3.3) and (2.3.4), for all $x \in \mathbb{R}^d$, $t \geq 1$:

$$\int_{\mathbb{R}^d} |\phi(x'')|^{1 + \delta} \int_{\mathbb{R}^d} (p^n_1(x', x''))^{1 + \delta} |\partial^n_x p^n_1(x, x')|^{1 + \delta} d x' d x'' \leq (C) \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\phi(x'')|^{1 + \delta} d x'' \leq C \int_{\mathbb{R}^d} \left( \int_{\mathbb{R}^d} (1 + |W(x'')|^{1 + \delta})^{1/m} e^{-c\frac{|x''|^2}{(x''\delta)^2}} d x'' \right) e^{-c|x-x''|^2} d x' < \infty.$$ 

If we intend to obtain the convergence to invariant measure under monotonic condition on the coefficients, i.e. we consider bounded, monotonic coefficients (Assumption 2.3.1 and 2.3.3 combined with $W(x) = |x|^2$), the typical example is constant $\Sigma$ and $\beta(s, x) = -xe^{-x}$ for all $(s, x) \in (0, \infty) \times \mathbb{R}^d$. Notice however that since we are bounding the coefficients $\beta, \Sigma$ only when restricted to $[0, 1]$, outside the interval $[0, 1]$ the growth can be arbitrary, allowing a broader selection of examples.

An alternative is the result presented by Eidelman in [26, Theorem VI.5].

**Lemma 2.3.8.** Suppose Assumptions 2.3.2 and 2.3.1(b) hold. Assume moreover that $W$ satisfies that

$$\sup_{c>0} \left\{ \int_{\mathbb{R}^d} |W(x)| e^{-c|x|^2} d x \right\} < \infty.$$ 

Then, given any $x \mapsto g(x) \leq C(1 + W(x))^{(1+m)/m}$, $C > 0$ and $\phi \in \mathcal{S}_m$ defined in (2.3.1) for some $m \in \mathbb{N}$, Assumptions 2.2.1 and 2.2.3 are satisfied for some $C, c, \delta > 0$ with $h(x) = Ce^{c|x|^2}$, for $x \in \mathbb{R}^d$.

**Proof.** Let $\beta$ and $\Sigma$ be the coefficients of the SDE (2.2.1) restricted to the time interval $[0, 1]$. Under Assumptions 2.3.2 and 2.3.1(b) by [26, Theorem VI.5], the fundamental solution $\Gamma$ to the PDE $(L - \partial_t) \Gamma = 0$ satisfies the following bounds for some $C, c$, $\nu > 0$ and for $n = 0, 1, 2$:

$$|\partial^n_x \Gamma(1, x''; 0, x)| \leq C \exp(-c|x - x''|^2 + \nu|x|^2 - \nu|x''|^2).$$

(2.3.5)

Closer to what we are looking for is the fundamental solution to the adjoint equation (2.3.2), which is denoted by $\Gamma^*$ and one time reversal away from $p^n_1(x, x'')$. Namely, $p^n_1(x, x'') = \Gamma^*(s - \tau, x''; \tau, \xi) = \Gamma_0(\tau, \xi; s - \tau, x'')$, where the last equality holds by the result proved in [26, Theorem VI.2] under the additional assumption that the coefficients are Hölder-continuous in space uniformly in time.

We can therefore conclude that Assumption 2.2.3 holds for functions $\gamma$ as in the statement of the lemma since by Young’s inequality:

$$\int_{\mathbb{R}^d} g(x'') |\partial^n_x p^n_1(x, x'')| d x'' \leq C \int_{\mathbb{R}^d} (1 + W(x''))^{(1+m)/m} e^{-c|x - x''|^2 + \nu|x|^2 - \nu|x''|^2} d x''$$

41
\[
    \leq C e^{v|x|^2} \left( \int_{\mathbb{R}^d} (1 + W(x''))^{1/m} e^{-v|x''|^2} e^{c|x|^2 - c/2|x''|^2} \, dx'' \right)
\]
\[
    \leq C e^{(v+c)|x|^2}.
\]

And, in fact if \( W \in B_p(\mathbb{R}^d) \), then
\[
    \int_{\mathbb{R}^d} (1 + W(x''))^{1/m} e^{-c|x''|^2} \, dx'' \leq C \int_{\mathbb{R}^d} (1 + |x''|^{p/m}) e^{-c|x''|^2} \, dx'' \leq C.
\]

We turn now to Assumption 2.2.1 and apply first Chapman–Kolmogorov’s identity, Leibniz formula and Jensen’s inequality. And again, by the bounds of the derivatives at \( t = 1 \) and the density at any \( t \geq 1 \) (see \[26\]), there exists a constant \( C_t > 0 \) dependent on \( t \) such that \( \sup p_t^1(x, x'') \leq C_t \). Finally, given the definition of \( \mathcal{S}_m \), there exist \( C > 0 \) such that \( \phi \leq C(1 + W)^{1/m} \). All together, we have that for all \( t \geq 1, x \in \mathbb{R}^d \) and by choosing \( \delta > 0 \), there exist \( C, C_t > 0 \) (where again we allow these constants to change from line to line in order to avoid cumbersome notation) such that
\[
    \int_{\mathbb{R}^d} |\phi(x'')| \partial_t p_t^0(x, x'') |1 + \delta \, dx''
\]
\[
    \leq \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} |\phi(x'')| p_t^1(x', x'') |1 + \delta \, | \partial_t p_t^0(x, x') |1 + \delta \, dx' \, dx''
\]
\[
    \leq C_t \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (1 + W(x''))^{(1+\delta)m} e^{-v|x'|^2 + e^{-c|x'|^2} - v|x''|^2} \, dx'dx''
\]
\[
    \leq C_t \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} e^{v|x|^2} (1 + W(x''))^{(1+\delta)m} e^{-v|x''|^2} e^{-c|x'|^2} \, dx' \, dx''
\]
\[
    \leq C_t \int_{\mathbb{R}^d} e^{(v+c)|x|^2} e^{-\delta|x|} (1 + \delta) c|x|^2 + (1 + \delta)|x||x''| - \delta v|x|^2 \, dx'
\]
\[
    \leq C_t e^{v(1+\delta)|x|^2} \left( \int_{\mathbb{R}^d} e^{v+c-\delta v}|x|^2 e^{-\delta(1+\delta)|x|} \, dx \right)
\]
\[
    \leq C_t e^{v(1+\delta)|x|^2} \left( \int_{\mathbb{R}^d} e^{v+c-\delta v}|x|^2 e^{(1+\delta)|x|^2} \, dx \right)
\]
\[
    \leq C_t e^{(v+c)(1+\delta)|x|^2} < \infty.
\]

Note that we could have formulated yet another assumption as an alternative to Assumption 2.3.1(a) and Assumption 2.3.1(b) by employing the result presented by Deck and Kruse [68, Corollary 4.2]. Suppose that Assumption 2.3.2 holds in addition of \( \beta, \Sigma \) belonging to \( C((0, 1] \times \mathbb{R}^d) \), being locally Hölder continuous in both components and globally Hölder growing in space (i.e. in \( B_p(\mathbb{R}^d) \) with \( p \in (0, 1) \)), uniformly in time for the interval \([0, 1]\). Deck and Kruse’s result allows us to conclude the following type of bounds for the transition density and its derivatives: for some \( h, \lambda^* > 0 \), \( c(h) > 0 \), \( 0 \leq n \leq 2 \), \( |\partial_x^n p_t^1(x, y)| \leq c(h) e^{h|y|^2} \exp\left( -\frac{\lambda^*|x-y|^2}{2(t-s)} \right) (t-s)^{1/2} \).

We omit the detailed computations since two illustrative examples were already presented and having a detailed catalogue is beyond the scope of this chapter.
2.3.2 Convergence to a measure under Lyapunov condition

Consider $\mathcal{F}_m$, the family of test functions defined in (2.3.1) and suppose that the Lyapunov function is of polynomial order. Since the increment of the elements in this family is controlled by the Lyapunov function itself, we prove in Theorem 2.3.10 that convergence to a static measure in Wasserstein distance (see Lemma 2.3.9) is enough to guarantee Assumption 2.2.2.

Lemma 2.3.9. Let Assumption 2.3.3 hold, $\Sigma$ be bounded and $\xi$ a random variable such that $\mathcal{L}(\xi) \in \mathcal{B}^{W^2}(\mathbb{R}^d)$. Assume moreover that there exists $c_W > 0$ such that $c_W|x|^p \leq W(x)$, for all $x \in \mathbb{R}^d$. Then there exists a measure $q \in \mathcal{B}^{W^2}(\mathbb{R}^d)$ and a constant $C > 0$ such that the density $x'' \mapsto p_{1+s}(x, x'')$ of the random variable $X_{1+s}$ defined by (2.2.1) satisfies the following:

$$\sup_{\phi \in \mathcal{F}_m} \left| \int_{\mathbb{R}^d} \phi(x') \left( p_{1+s}(x, dx') - q(dx') \right) \right| \leq \left( e^{-M_1s} \int_{\mathbb{R}^d} W(x-x'') q(dx'') \right)^{1/m}, \quad \forall (s, x) \in (0, \infty) \times \mathbb{R}^d. \tag{2.3.6}$$

The following proof of existence to an invariant measure, based on existence of Lyapunov functions and an adaptation of the underlying metrics to the studied problem, is not a new approach. It has been employed for example in [33, Corollary 2.1 and Remark 2.3] to prove convergence to invariant measure first in Total Variation distance and afterwards in Wasserstein distance.

Proof Lemma 2.3.9. Let $\lambda$ be an arbitrary positive constant and $\xi, \xi'$ two independent, distinct starting random variables $\mathcal{B}^{W^2}$. When Itô’s formula is applied to $W(X_{1,\xi} - X_{1,\xi'})$, we obtain:

$$d(e^{As}W(X_{1,\xi} - X_{1,\xi'}))$$

$$= e^{As} \left( \lambda W(X_{1,\xi} - X_{1,\xi'}) + (\partial_x W(X_{1,\xi} - X_{1,\xi'}))(\beta(s, X_{1,\xi}) - \beta(s, X_{1,\xi'})) \right)$$

$$+ \frac{1}{2} \text{tr} \left( (\Sigma^*(s, X_{1,\xi}) - \Sigma^*(s, X_{1,\xi'})) (\partial^2_x W(X_{1,\xi} - X_{1,\xi'})) (\Sigma(s, X_{1,\xi}) - \Sigma(s, X_{1,\xi'})) \right) ds$$

$$+ e^{As} \left( \partial_x W(X_{1,\xi} - X_{1,\xi'})(\Sigma(s, X_{1,\xi}) - \Sigma(s, X_{1,\xi'})) \right) dB_s.$$ 

Let us assume for now, and we will show it later in the proof, that for any $t > 1$

$$\mathcal{V}_{1,t} := \int_1^t e^{As} \left( \partial_x W(X_{1,\xi} - X_{1,\xi'})(\Sigma(s, X_{1,\xi}) - \Sigma(s, X_{1,\xi'})) \right) dB_s \tag{2.3.7}$$

is a martingale. In particular, it has 0 expectation. However, we can proceed assuming it for now and we will conclude the proof with a necessary stopping time argument. We continue by taking expectations of the above and using Assumption 2.3.3 to arrive to

$$d(e^{As}E[W(X_{1,\xi} - X_{1,\xi'})])$$

$$= e^{As} \left( \lambda E[W(X_{1,\xi} - X_{1,\xi'})] + E[(L(s, X_{1,\xi}) - L(s, X_{1,\xi'})) W(X_{1,\xi} - X_{1,\xi'})] \right) ds$$

$$\leq e^{As}(\lambda - M_1) E[W(X_{1,\xi} - X_{1,\xi'})] ds. \tag{2.3.8}$$

Therefore, for the particular choice of $\lambda = M_1 > 0$, the above means $d(e^{As}E[W(X_{1,\xi} - X_{1,\xi'})]) \leq$
\[
0 \, ds, \text{ which by integrating in between 1 and } s+1 > 1 \text{ implies that }
\]
\[
\mathbb{E}[W(X_{1+s}^{1,\xi} - X_{1+s}^{1,\xi'})] \leq e^{-M_s} \mathbb{E}[W(\xi - \xi')]. \quad (2.3.9)
\]

Next recall that we assumed the existence of \( c_W > 0 \) such that \( c_W |x|^p \leq W(x), \forall x \in \mathbb{R}^d \). Moreover, by the properties of \( W \) and the Wasserstein distance we have the following bound from \((2.3.9):
\]
\[
c_W W^p_p(\mathcal{L}(X_{1+s}^{1,\xi}), \mathcal{L}(X_{1+s}^{1,\xi'})) \leq c_W \mathbb{E}[|X_{1+s}^{1,\xi} - X_{1+s}^{1,\xi'}|^p] \leq \mathbb{E}[W(X_{1+s}^{1,\xi} - X_{1+s}^{1,\xi'})] \leq e^{-M_s} \mathbb{E}[W(\xi - \xi')] \leq e^{-M_s}(1 + \mathbb{E}[|\xi - \xi'|^p]) \leq C e^{-M_s} ||\xi - \xi'||^p. \quad (2.3.10)
\]

Notice next that inequality \((2.3.10)\) was obtained for arbitrary \( \xi, \xi' \) with certain distributions. Namely,
\[
c_W W^p_p(\mathcal{L}(X_{1+s}^{1,\xi}), \mathcal{L}(X_{1+s}^{1,\xi'})) \leq C e^{-M_s} ||\xi - \xi'||^p \quad (2.3.11)
\]
does not depend on the random variables themselves but rather their laws. We can therefore take the infimum over all couplings \((\xi, \xi')\) and on the right hand side of \((2.3.11)\) obtain:
\[
c_W W^p_p(\mathcal{L}(X_{1+s}^{1,\xi}), \mathcal{L}(X_{1+s}^{1,\xi'})) \leq C e^{-M_s} W^p_p(\mathcal{L}(\xi), \mathcal{L}(\xi')).
\]

We conclude:
\[
W_p(\mathcal{L}(X_{1+s}^{1,\xi}), \mathcal{L}(X_{1+s}^{1,\xi'})) \leq \frac{C^{1/p} e^{-M_s}}{c_W^{1/p}} W_p(\mathcal{L}(\xi), \mathcal{L}(\xi')). \quad (2.3.12)
\]

In particular, this bound implies convergence to invariant measure for the process \( X^{1,\xi}_{\cdot} \) given that \( \mathcal{L}(\xi) \in \mathcal{P}^W(\mathbb{R}^d) \). Indeed, first notice that the space \( \mathcal{P}^W(\mathbb{R}^d) \) equipped with the \( p \)-Wasserstein distance is a closed subspace (subspace because of the polynomial growth and properties of the increments of \( W \); and closed by definition of \( \mathcal{P}^W \) and polynomial growth of \( W \)) of the complete metric space of the probability measures with finite \( p \)-moments and the \( p \)-Wasserstein distance (see for e.g. [71]). Consequently, \((\mathcal{P}^W(\mathbb{R}^d), W_p)\) is a complete metric space itself. Moreover, for \( \delta > 0 \) large enough such that \( \frac{C e^{-M_1}}{c_W} < 1 \), inequality \((2.3.12)\) implies that the map \( \mathcal{F}_{\delta+1}(\mathcal{L}(\xi)) := \mathcal{L}(X_{\delta+1}^{1,\xi}) \) defined on this space is a contraction. Therefore the Banach Fixed Point Theorem asserts the existence of a unique fixed point \( \hat{q} \) for \( \mathcal{F}_\delta \), which at this stage could be dependent on \( \delta \).

However, one can construct another measure \( q := \frac{1}{\delta} \int_0^\delta \mathcal{F}_{\delta+1}(\hat{q}) \, ds \) which, by a similar argument to that in [72, Section 3], we next prove to be an invariant measure for \( \mathcal{F}_r \) with arbitrary \( r \in [0, \infty) \). Indeed,
\[
\mathcal{F}_r q = \frac{1}{\delta} \int_0^\delta \mathcal{F}_{\delta+1}(\hat{q}) \, ds = \frac{1}{\delta} \int_0^\delta \mathcal{F}_r \mathcal{F}_{\delta+1}(\hat{q}) \, ds = \frac{1}{\delta} \int_0^\delta \mathcal{F}_{\delta+1+r}(\hat{q}) \, ds = \frac{1}{\delta} \int_r^{\delta+r} \mathcal{F}_{\delta+1}(\hat{q}) \, ds = \frac{1}{\delta} \int_r^{\delta+r} \mathcal{F}_{\delta+1}(\hat{q}) \, ds + \frac{1}{\delta} \int_0^{\delta} \mathcal{F}_{\delta+1}(\hat{q}) \, ds = \frac{1}{\delta} \int_0^{\delta} \mathcal{F}_{\delta+1}(\hat{q}) \, ds = q.
\]
The uniqueness of such an invariant measure follows from (2.3.12). Indeed, suppose that \( \xi \) and \( \xi' \) are distributed accordingly to two different invariant measures \( q \) and \( q' \), respectively. Then, from (2.3.12)

\[
W_p(q, q') = W_p(\mathcal{L}(X_{1+s}^\xi), \mathcal{L}(X_{1+s}^{\xi'})) \leq C e^{-\frac{M_1}{p} s} W_p(q', q),
\]

and since this must hold for all \( s \geq 0 \), we arrive to the contradiction that \( q = q' \), concluding that way uniqueness of the invariant measure.

Moreover, if we choose \( \xi' \) distributed as the invariant measure denoted by \( q \) and we take test functions \( \phi \) belonging to \( \mathcal{S}_m \) defined in (2.3.1), display (2.3.9) implies after using Hölder’s inequality that:

\[
\begin{align*}
\sup_{\phi \in \mathcal{S}_m} \int_{\mathbb{R}^d} \phi(x'')(p_{1+s}(x, dx') - q(dx'')) &= \sup_{\phi \in \mathcal{S}_m} \int_{\mathbb{R}^d} \phi(x') p_{1+s}(x, dx') - \int_{\mathbb{R}^d} \phi(x'') q(dx'') \\
&= \sup_{\phi \in \mathcal{S}_m} \left| \mathbb{E}[\phi(X_{1+s}^1)] - \mathbb{E}[\phi(X_{1+s}^{1'})] \right| \\
&\leq \sup_{\phi \in \mathcal{S}_m} \mathbb{E}[\left| \phi(X_{1+s}^1) - \phi(X_{1+s}^{1'}) \right|] \leq \sup_{\phi \in \mathcal{S}_m} \left( \mathbb{E}[\left| \phi(X_{1+s}^1) - \phi(X_{1+s}^{1'}) \right|^{1/2}] \right)^2 \\
&\leq \left( \mathbb{E}[W(X_{1+s}^1 - X_{1+s}^{1'}))] \right)^{1/m} \leq e^{-\frac{M_1}{m} s} \left( \mathbb{E}[W(x - \xi')]) \right)^{1/m} \\
&= e^{-\frac{M_1}{m} s} \left( \int_{\mathbb{R}^d} W(x - x'') q(dx'') \right)^{1/m}.
\end{align*}
\]

Hence, we proved that (2.3.6) holds.

We conclude the proof by illustrating the stopping time argument. Let us define the following increasing sequence of stopping times: \( T_n := \inf \{t \geq 1 : |X_t^\xi| > n \} \rightarrow \infty, n \rightarrow \infty \). Notice that the stopped version of the integral process (2.3.7) is a uniformly integrable martingale. For the process in (2.3.7) to be a true martingale it is enough for the integrand to be square integrable. Since on one hand \( \Sigma \) is bounded and on the other hand there exists \( M_W > 0 \) such that \( |\partial_x W(x)| \leq M_W (1 + W(x)), \forall x \in \mathbb{R}^d \) (see Assumption 2.3.3) we get for all \( t > 1 \):

\[
\begin{align*}
\mathbb{E}[\int_1^{t \wedge T_n} (e^{\lambda s} \partial_x W(X_s^1, X_s^{1'}) - \Sigma(s, X_s^1, X_s^{1'})) (\Sigma(s, X_s^1, X_s^{1'}) - \Sigma(s, X_s^{1'})) ds] &\leq 2M^2 \mathbb{E}[\int_1^{t \wedge T_n} e^{2\lambda s} |\partial_x W(X_s^1, X_s^{1'}) - X_s^{1'})|^2 ds] \\
&\leq 2M^2 \mathbb{E}[\int_1^{t \wedge T_n} (1 + W(X_s^1, X_s^{1'})^2) ds].
\end{align*}
\]

Moreover, the stopped process defined from (2.3.7) is a martingale and when applied to it, inequality (2.3.9) (which also holds for \( W^2 \)) implies

\[
2M^2 \mathbb{E}[\int_1^{t \wedge T_n} (1 + W(X_s^1, X_s^{1'})^2) ds] \leq C \mathbb{E}[\int_1^{t \wedge T_n} e^{2\lambda s} (1 + W(\xi - \xi'))^2 ds].
\]

Given this uniform in \( n \) boundedness of the right hand side, when letting \( n \rightarrow \infty \) in the inequality above we conclude by Fatou's Lemma (allowing us to take the limit inside the
expectation on the left hand side) that
\[ 
\mathbb{E}\left[ \int_1^t \left( e^{\lambda s} \partial_s W(X_s^{1,\xi} - X_s^{1,\xi'}) (\Sigma(s, X_s^{1,\xi}) - \Sigma(s, X_s^{1,\xi'})) \right)^2 ds \right] 
\leq \liminf_{n \to -\infty} \left[ \int_1^{t \wedge T_n} \left( e^{\lambda s} \partial_s W(X_s^{1,\xi} - X_s^{1,\xi'}) (\Sigma(s, X_s^{1,\xi}) - \Sigma(s, X_s^{1,\xi'})) \right)^2 ds \right] 
\leq CE[W^2(\xi - \xi')] \int_1^t e^{2\lambda s} ds. 
\]
Finally, notice that since \( \mathbb{E}[W^2(\xi)] < \infty, \mathbb{E}[W^2(\xi')] < \infty \) and \( \xi, \xi' \) are independent, \( \mathbb{E}[W^2(\xi - \xi')] < \infty \). Hence, the stochastic integral in (2.3.7) is a martingale. \( \square \)

### 2.3.3 Result for autonomous SDEs

We are now ready to combine the results from the previous sections and apply Theorem 2.2.4 to the example falling under the set of Assumptions 2.3.1–2.3.2–2.3.3.

**Theorem 2.3.10.** Suppose that Assumptions 2.3.2, 2.3.3 and 2.3.1(a) or 2.3.1(b) hold. Assume moreover that there exists \( c_W > 0 \) such that \( c_W |x|^p \leq W(x), \forall x \in \mathbb{R}^d \) for the constant \( p \) in Assumption 2.3.3. Then for \( n = 1, 2 \) and for any \( \phi \in \mathcal{S}_m \) defined in (2.3.1):
\[
|\partial_x^n V(s, x)| \leq e^{(-M_1/m)s} h(x), \quad \text{for all } (s, x) \in (1, \infty) \times \mathbb{R}^d,
\]
where for some \( C, c > 0 \), we have
\[
h(x) = \begin{cases} 
Ce^{c|x|^2}, & \text{if } p < m, \\
C(1 + |x|^{p/m}), & \text{if } p \geq m; 
\end{cases}
\]
under Assumption 2.3.1(a) or \( h(x) = Ce^{c|x|^2} \) under Assumption 2.3.1(b).

**Proof.** By Lemma 2.3.9, we have seen that under Assumption 2.3.3, we have
\[
\sup_{\phi \in \mathcal{S}_1} \left| \int_{\mathbb{R}^d} \phi(x')(p_1^1(x', d x') - q(d x')) \right| \leq e^{-(M_1/m)s} \left( \int_{\mathbb{R}^d} W(x - x') q(d x') \right)^{1/m}.
\]
And since
\[
\int_{\mathbb{R}^d} W(x - x'') q(d x'') \leq \int_{\mathbb{R}^d} C(1 + |x - x''|^{p/m}) q(d x'') \leq C(1 + |x|^p),
\]
we know that Assumption 2.2.2 is satisfied with \( g(x) = C(1 + |x|^{p/m}) \) and \( G(s) = e^{-(M_1/m)s} \).

Moreover, under Assumptions 2.3.2 and 2.3.1(a), recall that Lemma 2.3.7 asserts that Assumption 2.2.3 is satisfied with \( h(x) = Ce^{c|x|^2} \) since by Young’s inequality:
\[
\int_{\mathbb{R}^d} (1 + |x''|^{p/m}) Ce^{-c|x - x''|^2} d x'' \leq \int_{\mathbb{R}^d} (1 + |x''|^{p/m}) Ce^{-c|x|^2 - c|x''|^2 + c|x||x''|} d x'' 
\leq \int_{\mathbb{R}^d} (1 + |x''|^{p/m}) Ce^{c|x|^2 - c/2|x''|^2} d x'' 
\leq Ce^{c|x|^2}.
\]

46
Now, if \( p \geq m \) by a mere change of variables, we conclude on a similar fashion to Lemma 2.3.7 that the bound is achieved with \( h(x) = C(1 + |x|^{p/m}). \)

Alternatively, under Assumptions 2.3.2 and 2.3.1(b), in Lemma 2.3.8 we prove that Assumption 2.2.3(i) is satisfied with \( h(x) = C e^{(v+c)|x|^2}. \)

Moreover, Assumption 2.2.1 is proved under both Assumptions 2.3.1(a) and 2.3.1(b) again in Lemmas 2.3.7 and 2.3.8 respectively.

Finally, since Assumptions 2.2.2, 2.2.1 and 2.2.3 are satisfied, Theorem 2.2.4 gives us the claim.

**Monotonic case**

In this section we present a more restrictive, although also more straightforward, set of assumptions for guaranteeing the conclusion of Theorem 2.3.10. Instead of the abstract Lyapunov functions we work with in Theorem 2.3.10, we work directly with assumptions for guaranteeing the conclusion of Theorem 2.3.10. Instead of the abstract order locally Lipschitz functions:

\[
\mathcal{S}_m := \{ \phi : \mathbb{R}^d \rightarrow \mathbb{R} \mid \phi \in B(\mathbb{R}^d) \text{ and } \exists C > 0 \text{ s.t } |\phi(x) - \phi(y)| \leq C(1 + |x|^{m/2} + |y|^{m/2})|x - y|, \forall x, y \in \mathbb{R}^d \}.
\]

(2.3.13)

The monotonicity and locally Lipschitz conditions are popular assumptions in the field and we see next that this example is a special case of that presented in this section. This means that, as a particular case of Theorem 2.3.10, the same type of decaying in time estimates hold for this scenario.

**Corollary 2.3.12.** Suppose that Assumptions 2.3.2, 2.3.11 and Assumption 2.3.1(a) or 2.3.1(b) hold. Then for \( n = 1, 2 \) and for any \( \phi \in \mathcal{S}_m \) defined in (2.3.13) for \( m \in \mathbb{N} \) such that \( p/m \geq 2 \), the following holds:

\[
|\partial^p_x V(s,x)| \leq e^{-M t} h(x), \quad \forall (s,x) \in (1, \infty) \times \mathbb{R}^d;
\]

where for some \( C > 0 \), \( h(x) = C(1 + |x|^{m}) \) under Assumption 2.3.1(a) or \( h(x) = C(1 + |x|^{m})e^{v|x|^2} \) under Assumption 2.3.1(b).

**Proof.** First notice that the monotonicity Assumption 2.3.11 coincides reduces to the Lyapunov Assumption 2.3.3 whenever \( W(x) = |x|^p \). Moreover, for \( \phi \in \mathcal{S}_m \) with \( \frac{p}{m} \geq 2 \),

\[
|\phi(x) - \phi(y)| \leq |x - y|^{\frac{p}{m}} = |x - y|^{2} |x - y|^{\frac{p}{m} - 2} \leq (|x| + |y|)|x - y| |x - y|^{\frac{p}{m} - 2}
\]

\[
\leq \ldots \leq (|x|^{\frac{p}{m} - 1} + |y|^{\frac{p}{m} - 1})|x - y|,
\]

i.e \( \phi \in \mathcal{S}_m \) and \( W(x) = |x|^p \), implies \( \phi \in \mathcal{S}_m \). This and Theorem 2.3.10 give the claim.

\[ \square \]
Let us conclude with a few remarks on the relation in between the the family of test functions \( S_m \) and \( S'_m \). First, the family \( S_m \) is significantly wider than that of \( S'_m \). For one, if \( W \) is a polynomial of order \( p \), the ratio in between \( p \) and \( m \) will determine if the admissible test functions in \( S_m \) are either locally Lipschitz or locally Hölder, while the functions in \( S'_m \) must be locally Lipschitz. If instead \( W \) is polynomially growing only, then the test functions in \( S_m \) can have additionally a bounded component added to this locally Lipschitz or locally Hölder part.

Moreover, although the locally Lipschitz condition seems more natural than the one presented through the Lyapunov function, it doesn't give itself easily to generalization. Indeed, the family \( S_m \) can used with other than polynomial Lyapunov functions as long as one can obtain contraction in \( W \)-Weighted Total Variation distance and \( W \) is still integrable against the bounds obtained for the derivatives of the transition density (see Section 2.4 below).

### 2.4 Application to non–autonomous SDE decaying to an autonomous one

Consider again the non–autonomous SDE (2.2.1). Given the time–varying character of the coefficients, these processes do not easily have an invariant measure. Recall however that Assumption 2.2.2 was not asking for one, just for a static measure to which the transition probabilities “stick”.

Additional intuition behind the result in Theorem 2.2.4 could be extracted when noticing that if there exist limiting functions \( \beta_\infty(x) := \lim_{r \to \infty} \beta(x, r) \) and \( \sigma_\infty(x) := \lim_{r \to \infty} \Sigma(x, r) \), we can define the auxiliary process \( Z \) as a solution to the widely studied autonomous SDE:

\[
d Z^T_x = \beta_\infty(Z^T_s, x) \, ds + \sigma_\infty(Z^T_s, x) \, dB_s; \quad Z^T_\tau, x = x. \tag{2.4.1}
\]

We will formulate conditions so that \( Z \) has an invariant measure \( q \) which, given its autonomous character, are quite lax. Moreover, we will show that the transition densities of the process \( X \) decay to \( q \). This together with the same regularity of \( \beta, \Sigma \) as in Section 2.3 will let us verify Assumptions 2.2.3 and 2.2.2 and conclude the decay of the derivatives in space of \( V(t, x) = \mathbb{E}[\phi(X^t_\cdot)] \). Moreover, unlike in Section 2.3, the test functions \( \phi \) are not restricted to polynomial growth but are rather controlled by a Lyapunov function which we specify later.

In order to do so, we can no longer use the Wasserstein metric to obtain convergence to invariant measure of \( Z \). The Weighted Total Variation (WTV) distance (see (1.7.1) for the definition and [3, 73, Theorem 8.9] for more detail) proves to be more appropriate in this setting. The next two remarks consider the relation in between this distance and the TV and Wasserstein distances.

**Remark 2.4.1.** For any \( W : \mathbb{R}^d \to \mathbb{R}_+ \) measurable, \( (\mathcal{P}^W(\mathbb{R}^d), || \cdot ||_{TV}) \subseteq (\mathcal{P}(\mathbb{R}^d), || \cdot ||_{TV}) \) and for any \( q, q' \in \mathcal{P}^W(\mathbb{R}^d) \) we have that

\[
2 || q - q' ||_{TV} \leq \int_{\mathbb{R}^d} (1 + W(x)) | q - q' | (dx) 
= \left( \int_{\mathbb{R}^d} (1 + W(x)) q(dx) \right) \int_{\mathbb{R}^d} \left| \frac{(1 + W(x))}{\int_{\mathbb{R}^d} (1 + W(x)) q(dx)} q - \frac{(1 + W(x))}{\int_{\mathbb{R}^d} (1 + W(x)) q(dx)} q' \right| (dx).
\]
We can then define new probability measures $Q(dx) = \frac{(1+W(x))}{\int_{\mathbb{R}^d}(1+W(x))q(dx)}q(dx)$ and $Q'(dx) = \frac{(1+W(x))q'(dx)}{\int_{\mathbb{R}^d}(1+W(x))q(dx)}q(dx)$, which are absolutely continuous with respect to $q$ and $q'$ respectively. So, if we let $C_W = \frac{\int_{\mathbb{R}^d}(1+W(x))q'(dx)}{\int_{\mathbb{R}^d}(1+W(x))q(dx)}$,

$$2\|q-Q\|_{TV} \leq \|q-q'\|_W = \|q\|_W\|Q-C_WQ'\|_{TV}.$$ \hspace{1cm} (2.4.2)

Hence, by the first inequality, every Cauchy sequence in WTV is Cauchy in TV. Moreover, given a sequence $(q^n)_{n \in \mathbb{N}} \subset \mathcal{P}^W(\mathbb{R}^d)$ and if for any $n, m \in \mathbb{N}$ we assume $\|q^n - q^m\|_W \to 0$, by the reversed triangle inequality we know that as $n, m \to \infty$,

$$\left(\int_{\mathbb{R}^d}(1+W(x))q^m(dx)\right)/(\int_{\mathbb{R}^d}(1+W(x))q^n(dx)) \to 1 \text{ and } \int_{\mathbb{R}^d} \left| \frac{(1+W(x))}{\int_{\mathbb{R}^d}(1+W(x))q^m(dx)}q^n - \frac{(1+W(x))}{\int_{\mathbb{R}^d}(1+W(x))q^m(dx)}q^m \right|(dx) \to 0.$$

In particular, by (2.4.2), this means that every Cauchy sequence $(q^n)$ in $(\mathcal{P}^W(\mathbb{R}^d), \|\cdot\|_W)$ is also Cauchy in $(\mathcal{P}^W(\mathbb{R}^d), \|\cdot\|_{TV})$. And since $\|\cdot\|_{TV}$ is complete for the space of measures with finite mass, there must exist $q \in \mathcal{P}(\mathbb{R}^d)$ such that $q^n \to q$ as $n \to \infty$.

We conclude completeness of any closed convex space of measures in $\mathcal{P}^W(\mathbb{R}^d)$ with respect to $\|\cdot\|_W$. And in fact (see [31, Proof Theorem 3.2]), $(\mathcal{P}^W(\mathbb{R}^d), \|\cdot\|_W)$ is complete itself.

**Remark 2.4.2.** Note that for polynomial weight functions of degree $p$, it is equivalent to the $p-$Wasserstein metric (see [3, Theorem 6.15]). Indeed, if $W(x) = |x|^p$ for some $p \in \mathbb{N}$, where by [3, Theorem 6.15] we have that:

$$W^p_p(q,q') \leq 2^{p/(p-1)}\|q-q'\|_W.$$ 

In other words, the Weighted Total Variation distance controls the Wasserstein one for polyno-
mial functions $W$. Moreover, using the same trick as in Lemma 2.3.9, for polynomial weight
(i.e function $W$) of degree $p$, the opposite inequality holds up to a constant.

Let us quantify now the above qualitative statements. A possible set of assumptions, in
addition to Assumptions 2.3.1 and 2.3.2 is the following.

**Assumption 2.4.3** (Time dependence of the coefficients). There exist $\beta_\infty: \mathbb{R}^d \to \mathbb{R}^d$ and $\Sigma_\infty: \mathbb{R}^d \to \mathbb{R}^{d \times d}$ measurable such that

$$\beta_\infty(x) := \lim_{r \to \infty} \beta(x, r) \text{ and } \sigma_\infty(x) := \lim_{r \to \infty} \Sigma(x, r), \quad \forall (t, x) \in (0, \infty) \times \mathbb{R}^d.$$

Notice that as a consequence of this and Assumptions 2.3.1 and 2.3.2, there exists $L_\infty(x) := \lim_{r \to \infty} L(r, x), \forall x \in \mathbb{R}^d$ which also inherits regularity and uniform ellipticity from $L$.

**Assumption 2.4.4** (Lyapunov function). There exists a function $W: \mathbb{R}^d \to [0, \infty)$ such that

$$\sup_{x \to \infty} |\beta(t, x)| \leq M_2 W(x) \text{ and } \sup_{x \to \infty} |\Sigma(t, x)| \leq M_2 W(x), \quad \forall x \in \mathbb{R}^d.$$ \hspace{1cm} (2.4.3)
Additionally, we assume $W \in C^2(\mathbb{R}^d)$ and there exists a constant $M_1 > 0$ such that for all $t \geq 0$ and $x \in \mathbb{R}^d$,

$$L(t, x)W(x) \leq -M_1 W(x). \quad (2.4.4)$$

We can see that (2.4.3) and (2.4.4) also hold with $\beta, \Sigma, L$ replaced by $\beta_\infty, \Sigma_\infty, L_\infty$. Hence, (2.4.1) has a unique weak solution by [74, Theorem 2.4].

Moreover, we work with the family of test functions:

$$\mathcal{S} := \left\{ \phi : \mathbb{R}^d \to \mathbb{R} | \phi \in B(\mathbb{R}^d) \text{ and } \sup_{x \in \mathbb{R}^d} \frac{|\phi(x)|}{W(x)} < \infty \right\}. \quad (2.4.5)$$

**Remark 2.4.5.** Although this set of assumptions might seem restrictive at first sight, it is more general than what is available in the current literature. Indeed, in addition to the generalization to the non-autonomous dynamics, when compared to [27], we notice one extra advantage of the main results of this section, Theorem 2.4.8: there is no need for smoothness and boundedness of all derivatives for the coefficients. In other words, even when discussing the time decay of the space derivatives of $(t, x) \mapsto \mathbb{E}[\phi(Z_{t}^{0, x})]$, our results hold under weaker assumptions on the coefficients $\beta_\infty, \Sigma_\infty$.

Next we state a technical lemma which gives a priori uniform estimates ensuring well-posedness and is used to compare the auxiliary autonomous process $Z$ to the original non-autonomous process $X$.

**Lemma 2.4.6.** Let Assumption 2.4.4 hold and $\Sigma$ be bounded. Then, for all $\xi$ such that $\mathbb{E}[W(\xi)] < \infty$ and all $\tau \geq 1$, the solutions $(X_{t+s}^{T, \xi})_{s \geq 0}$ and $(Z_{t+s}^{T, \xi})_{s \geq 0}$ to (2.2.1) and (2.4.1) respectively, satisfy the following:

$$\mathbb{E}[W(X_{t+s}^{T, \xi})] \leq e^{-M_1 s} \mathbb{E}[W(\xi)] \quad \text{and} \quad \mathbb{E}[W(Z_{t+s}^{T, \xi})] \leq e^{-M_1 s} \mathbb{E}[W(\xi)], \quad \forall s > 0.$$ 

**Proof.** An analogous computation to Lemma 2.3.9, applying Itô’s formula to the process $(e^{\lambda s}W(X_{t+s}^{T, \xi}))_{s \geq 0}$ for arbitrary $\lambda \in \mathbb{R}$ and using (2.4.4), leads to:

$$d\left(e^{\lambda s} \mathbb{E}[W(X_{t+s}^{T, \xi})]\right) \leq e^{\lambda s}(\lambda - M_1) \mathbb{E}[W(X_{t+s}^{T, \xi})] \, ds.$$ 

Hence, for $\lambda = M_1$ and for all $s > 0$, after integrating the previous inequality, we obtain the claimed bound:

$$\mathbb{E}[W(X_{t+s}^{T, \xi})] \leq \mathbb{E}[W(\xi)] e^{-M_1 s}.$$ 

Moreover, since as a consequence of Assumption 2.4.3, inequality (2.4.4) is also satisfied by the limiting generator $L_\infty$, we have the analogous is satisfied also by $Z$:

$$\mathbb{E}[W(Z_{t+s}^{T, \xi})] \leq \mathbb{E}[W(\xi)] e^{-M_1 s}.$$ 

$\Box$

Following [30], we prove in Proposition 2.4.7 below that under the Lyapunov Assumption 2.4.4, one obtains Assumption 2.2.2 by means of convergence in WTV. Recall that the Weighted Total Variation norm and distances, represented by $\| \cdot \|_W$ and $d_W$, are stated in (1.7.1).
**Proposition 2.4.7.** Let Assumptions 2.3.1, 2.3.2, 2.3.3, 2.4.3 and 2.4.4 hold and consider a random variable such that \( \mathcal{L}(\xi) \in \mathcal{D}_W(\mathbb{R}^d) \). Let us denote for any \( x \in \mathbb{R}^d \) the density of \((Z^{1,s}_{x,s})_{s \geq 0}\) (solution to (2.4.1)) by \( \mathbb{R}^d \ni x'' \mapsto p^{1+s}_s(x, x'') \). Then, \((Z^{1,\xi})_{s \geq 0}\) has an invariant measure \( q \in \mathcal{D}_W(\mathbb{R}^d) \) and the following is satisfied for some \( C, c > 0 \):

\[
d_W(\mathcal{L}(Z^{1,\xi}), q) \leq Ce^{-cs}d_W(\mathcal{L}(\xi), q). \tag{2.4.6}
\]

Moreover, for any \( \mathcal{S} \) defined by (2.4.5), there exist \( C, c > 0 \) such that

\[
\sup_{\phi \in \mathcal{S}} \left| \int_{\mathbb{R}^d} \phi(x')(p^{1+s}_s(x, dx') - q(dx')) \right| \leq Ce^{-cs} \int_{\mathbb{R}^d} W(x - x'')q(dx''), \ \forall (s, x) \in (0, \infty) \times \mathbb{R}^d. \tag{2.4.7}
\]

Noticing that \( W \geq 1 \), we conclude from this the final statement.

**Proof.**

**Step 1:** For the purposes of this proof let us introduce the semigroup \((\mathcal{T}_{1+s})_{s \geq 0}\) defined by the generator \( C^2(\mathbb{R}^d) \ni u \mapsto L_\infty u = \frac{1}{2} \text{tr} \left( \Sigma_\infty \Sigma_\infty \Sigma_\infty^2 u \right) + \beta_\infty \partial_x u \). Let us also assume for now that there exists an invariant measure \( q \in \mathcal{D}_W(\mathbb{R}^d) \), and we will prove this claim later. We know that \((\mathcal{T}_{1+s})_{s \geq 0}\) indeed exists, is unique and is a Markov semigroup on \( L^1(\mathcal{L}(\xi)) \) (see [8, Theorem 5.2.2, Proposition 5.2.5 and Example 5.5.1]). By [8, Theorem 6.4.7] we know that there exists a positive continuous function \( (t, x, x'') \rightarrow p_{1+t}(x, x'') \) such that for any \( \psi \in L^1(q) \) the following identity holds: \( \mathcal{T}_{1+t} \psi(x) = \int_{\mathbb{R}^d} p_{1+t}(x, x'') \psi(x'')dx'', \ \forall x \in \mathbb{R}^d \). Moreover, for any fixed \( x \in \mathbb{R}^d \), it satisfies \( \partial_t p_{1+t}(x, x'') = \mathcal{L}^*_\infty(x, x'')p_{1+t}(x, x'') \) for all \( t \geq 0, x'' \in \mathbb{R}^d \).

We know moreover that for a fixed \( x \in \mathbb{R}^d \), the density \( (s, x'') \rightarrow p^{1+s}_s(x, x'') \) satisfies the Fokker–Planck equation:

\[
\partial_t p^{1+s}_s(x, x'') = L^*_\infty(x, x'') p^{1+s}_s(x, x''); \quad p^1_s(x, x'') = \delta_x(x''), \ s \geq 0, x'' \in \mathbb{R}^d.
\]

By uniqueness of solution under our regularity assumptions on the coefficients (see [2, Theorem 7.4] or [25]) we know that \( p_{1+s}(x, x'') = p^{1+s}_1(x, x'') \) and therefore for any \( \psi \in L^1(q) \), \( \mathcal{T}_{1+t} \psi(x) = \int_{\mathbb{R}^d} p_{1+t}(x, x'') \psi(x'')dx'', \ \forall x \in \mathbb{R}^d, t \geq 0 \). Moreover, its dual is \( \mathcal{T}^*_{1+t} \sigma(x'') = \int_{\mathbb{R}^d} p_{1+t}(x, x'') \sigma(dx), \forall x'' \in \mathbb{R}^d, t \geq 0 \) and \( \sigma \in \mathcal{P}(\mathbb{R}^d) \).

Notice next that by Lemma 2.4.6 we know that for all \( (s, x) \in (0, \infty) \times \mathbb{R}^d: \)

\[
\mathcal{T}_{1+s} W(x) = \int_{\mathbb{R}^d} p^{1+s}_s(x, x'') W(x'')dx'' \leq W(x) e^{-M_1s}. \tag{2.4.8}
\]

**Step 2:** We will verify that we can apply Harris Ergodic Theorem.

First, due to (2.4.8), the function \( W : \mathbb{R}^d \rightarrow [0, \infty) \) and \( M_1 > 0 \) satisfy

\[
\int_{\mathbb{R}^d} W(y)p^{1+s}_s(x, dy) \leq e^{-M_1s} W(x), \ \forall (s, x) \in (0, \infty) \times \mathbb{R}^d. \tag{2.4.9}
\]

Second, recall that \( |\beta_\infty(x)| \leq \max_{|y| \leq |x|} W(y) \) for all \( y \in \mathbb{R}^d \). Then, for any fixed \( R > 0 \) and any fixed \( \hat{s} > 0 \), Harnack's inequality [8, Theorem 8.2.1] gives \( \kappa(\hat{s}) > 0 \) such that

\[
p^{1+\hat{s}}_1(x, x'') \geq \min_{|x| \leq R} p^{1+\hat{s}/2}_1(x, 0) \exp \left( -\kappa(\hat{s}) \left( 1 + \max_{|y| \leq |x''|} W(y) \right)^2 \right), \ \forall x'', \in \mathbb{R}^d, \ |x| \leq R. \tag{2.4.10}
\]

Let \( m(R) := \min_{|x| \leq R} p^{1+\hat{s}/2}_1(x, 0) \). Next we will show that \( m(R) > 0 \). Indeed, following [30, Proof
of Lemma 3.5], let us consider \( \psi \in C^\infty(\mathbb{R}^d) \) with compact support such that \( \psi(y) = 1 \) if \( |y| \leq 2R \) and \( \psi(y) = 0 \) if \( |y| > 3R \) and let \( |x| \leq R \). For \( t \geq 0 \) we see that:

\[
\int_{\mathbb{R}^d} \psi(y)p^1_{1+t}(x,dy) = \psi(x) + \int_0^t \int_{\mathbb{R}^d} L_\infty \psi(y)p^1_{1+s}(x,dy)ds.
\]

Consequently,

\[
\int_{\mathbb{R}^d} \psi(y)p^1_{1+t}(x,dy) \geq 1 - t\text{sup}_y|L_\infty \psi(y)|.
\]

Now by choosing \( t \) small, we have that \( \int_{\mathbb{R}^d} \psi(y)p^1_{1+t}(x,dy) \geq 1/2 \) for all \( |x| \leq R \). Since this holds for any such test function \( \psi \), we conclude that \( \min |p^1_{1+t}(x,y) \geq 1/2 \) for all \( |x| \leq R \). And choosing \( s \) smaller if necessary, we can assume that \( t = \frac{s}{4} \) and applying Harnack’s inequality again to prove that there exists \( C > 0 \) such that \( 1/2 \leq Cp^1_{1+\frac{s}{4}}(x,0) \) or, in other words, \( m(R) > 0 \) for every \( R > 0 \).

Notice then that by (2.4.10), there exists \( k \in (0,1) \) such that if \( R > 1/(1-e^{-M_1s}) \), then the probability measure

\[
\mu(dy) = \frac{1}{k}m(R)\exp\left(-\kappa(s)(1 + (\max_{|x^0| \leq 2|y|} W(x^0))^2 + |y|^2)\right)dy
\]

satisfies

\[
\inf_{x:W(x) \leq R} p^1_{1+s}(x,\cdot) \geq k\mu. \quad (2.4.11)
\]

Then, by (2.4.9) and (2.4.11), we can apply the Harris Ergodic Theorem [31, Theorem 1.3] and conclude the existence of two numbers, \( \tilde{\beta} \in (0,1) \) and \( \tilde{\beta} > 0 \), such that for every two probability measures on \( \mathbb{R}^d \), \( \nu_1, \nu_2 \),

\[
\|\int_{\mathbb{R}^d} p^1_{1+\frac{s}{2}}(x,\cdot)\mu_1(dx) - \int_{\mathbb{R}^d} p^1_{1+\frac{s}{2}}(x,\cdot)\mu_2(dx)\|_{\tilde{\beta}W} \leq \tilde{\beta}^n\|
u_1 - \nu_2\|_{\tilde{\beta}W}. \quad (2.4.12)
\]

Next, we apply repeatedly (2.4.12) for two initial laws \( \mathcal{L}(\xi) \) and \( \mathcal{L}(\xi') \) and \( (T_{1+nT}^*)_{n \geq 0} \). This allows us to conclude that for any \( n \in \mathbb{N} \) and fixed \( T > 0 \), there exist (some other) \( \tilde{\beta} > 0, \tilde{\beta} \in (0,1) \) such that

\[
||T_{1+nT}^*\mathcal{L}(\xi) - T_{1+nT}^*\mathcal{L}(\xi')||_{\tilde{\beta}W-1} \leq \tilde{\beta}^n||\mathcal{L}(\xi) - \mathcal{L}(\xi')||_{\tilde{\beta}W-1}.
\]

In particular, since

\[
||T_{1+nT}^*\mathcal{L}(\xi) - T_{1+nT}^*\mathcal{L}(\xi')||_{W}
\leq ||T_{1+nT}^*\mathcal{L}(\xi) - T_{1+nT}^*\mathcal{L}(\xi')||_{TV} + ||T_{1+nT}^*\mathcal{L}(\xi) - T_{1+nT}^*\mathcal{L}(\xi')||_{W}
\leq ||T_{1+nT}^*\mathcal{L}(\xi) - T_{1+nT}^*\mathcal{L}(\xi')||_{TV} + \tilde{\beta}^{-1}||T_{1+nT}^*\mathcal{L}(\xi) - T_{1+nT}^*\mathcal{L}(\xi')||_{\tilde{\beta}W-1}
\leq \max(1,\tilde{\beta}^{-1})||T_{1+nT}^*\mathcal{L}(\xi) - T_{1+nT}^*\mathcal{L}(\xi')||_{\tilde{\beta}W}
\leq \max(1,\tilde{\beta}^{-1})\tilde{\beta}^n||\mathcal{L}(\xi) - \mathcal{L}(\xi')||_{\tilde{\beta}W}
\leq \frac{\max(1,\tilde{\beta})}{\min(1,\tilde{\beta})}\tilde{\beta}^n||\mathcal{L}(\xi) - \mathcal{L}(\xi')||_{W},
\]

52
we conclude

\[ \|T_{1+nT}^s(\mathcal{L}(\xi)) - T_{1+nT}^s(\mathcal{L}(\xi'))\|_W \leq C\beta^n \|\mathcal{L}(\xi) - \mathcal{L}(\xi')\|_W. \] (2.4.13)

Moreover, for any function \(\psi \in C_0^\infty(\mathbb{R}^d)\) such that \(|\psi| \leq CW\) and \(0 \leq t < T\), we have \(|\mathcal{F}_t \psi| \leq \mathcal{F}_t |\psi| \leq C\mathcal{F}_t(W) \leq 2CW\). Hence,

\[ \int_{\mathbb{R}^d} \psi(x'')(T_{1+t}^s(\mathcal{L}(\xi)) - T_{1+t}^s(\mathcal{L}(\xi')))(x'') = \int_{\mathbb{R}^d} (\mathcal{F}_{1+t} \psi)(x'')(\mathcal{L}(\xi) - \mathcal{L}(\xi'))(dx'') \leq 2C\|\mathcal{L}(\xi) - \mathcal{L}(\xi')\|_W. \]

Since \(T > 0\) was arbitrary, this and (2.4.13) imply that there exist \(C, c > 0\) such that for all \(s > 0\),

\[ \|T_{1+s}^s(\mathcal{L}(\xi)) - T_{1+s}^s(\mathcal{L}(\xi'))\|_W \leq Ce^{-cs}\|\mathcal{L}(\xi) - \mathcal{L}(\xi')\|_W. \] (2.4.14)

Now, as long as we have existence of an invariant measure \(q\), applying this with \(\xi' \sim q\) allows us to conclude (2.4.6).

**Step 3:** Now, if we were to rely on the completeness \((\mathcal{P}W(\mathbb{R}^d), \|\cdot\|_W)\) stated in [31], we can use the contraction (2.4.6) (for \(s\) big enough such that \(Ce^{-cs} < 1\)) to carry out a similar argument to that presented in Proposition 2.3.9. Namely we obtain, by Banach Fixed Point theorem, the existence of an invariant measure \(q \in \mathcal{P}W(\mathbb{R}^d)\).

**Step 4:** Finally, we apply (2.4.6) with the initial condition \(\xi = x \in \mathbb{R}^d (a.s)\) and the fact that any \(\phi \in \mathcal{I}\) satisfies for some \(C > 0\) that \(\phi \leq C(1 + W)\), in order to obtain for some \(C, c > 0\):

\[
\sup_{\phi \in \mathcal{I}} \left| \int_{\mathbb{R}^d} \phi(x')(p_{1+s}^1(x, dx') - q(dx')) \right| \leq \sup_{\phi \in \mathcal{I}} \int_{\mathbb{R}^d} |\phi(x')| |p_{1+s}^1(x, dx') - q(dx')| \\
\leq Cd_W(p_{1+s}^1(x, \cdot), q(\cdot)) \leq Ce^{-cs}d_W(\delta_x, q) \\
= Ce^{-cs} \int_{\mathbb{R}^d} (1 + W(x - x'))q(dx'') .
\]

\[\square\]

As a final remark, let us mention that there are even weaker alternative Lyapunov conditions allowing us to conclude convergence in WTV. An option is assuming that there exists a function \(W(x) : \mathbb{R}^d \to [0, \infty)\), constants \(M_1 > 0, M_2 \in \mathbb{R}\) and a compact set \(K \subseteq \mathbb{R}^d\) with index function \(1_K\), such that for all \(t \in [0, \infty), x \in \mathbb{R}^d\),

\[ L(t, x)W(x) \leq M_2 1_K - M_1 W(x). \]

Then we have exponential decay to an invariant measure in Weighted Total Variation distance (see [73, Theorem 8.7]) for test functions in \(\mathcal{I}\) defined by (2.4.5).

Next, we state the main theorem of this section: another example of time decaying derivative estimates for \(V(t, x) = \mathbb{E}[\phi(X_t^x)]\), where \(X\) is solution to (2.2.1) and \(\phi \in \mathcal{I}\) defined by (2.4.5).

**Theorem 2.4.8.** Suppose that Assumptions 2.3.2, 2.3.1, 2.4.3 and 2.4.4 hold. Assume moreover that there exists \(C_W > 0\) such that \(\int_{\mathbb{R}^d} |W(x)|e^{-c|x|^2} \ dx \leq C_W\), where \(c > 0\) is any of the constants in (2.3.3), (2.3.4), (2.3.5). Then, there exists a measure \(q \in \mathcal{P}W(\mathbb{R}^d)\) such that, for every \(x \in \mathbb{R}^d\),

\[
\]
\( \mathbb{R}^d, s \geq 0, \) it is a static measure for \( Z_{1+s}^{1,x} \) solution to (2.4.1). Moreover, for \( n = 1, 2, \phi \in \mathcal{F} \) defined in (2.4.5), there exist \( C, c > 0 \) such that:

\[
|\frac{\partial}{\partial x} V(s, x)| \leq C e^{-cs} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (1 + W(x'' - x')) q(d x') e^{-c|x-x''|^2} d x'', \quad \text{for all } (s, x) \in (1, \infty) \times \mathbb{R}^d.
\]

**Proof.** Recall that for any \( x \in \mathbb{R}^d \) the density of \( (Z_{1+s}^{1,x})_{s \geq 0} \) (solution to (2.4.1)) is denoted by \( x' \mapsto p_{1+s}(x, x') \) and it has an invariant measure denoted by \( q \) by Proposition 2.4.7. Let the density of \( (X_{1+s}^{1,x})_{s \geq 0} \) (solution to (2.4)) be denoted by \( x' \mapsto P_{1+s}(x, x') \).

Then, first by the triangle inequality and afterwards by Lemma 2.4.6, for any \( x \in \mathbb{R}^d, s \geq 0, \)

\[
\int_{\mathbb{R}^d} W(x') |p_{1+s}^1(x, d x') - p_{1+s}^1(x, d x')| = \int_{\mathbb{R}^d} W(x') P_{1+s}(x, x') dx' + \int_{\mathbb{R}^d} W(x') p_{1+s}^1(x, x') dx' \leq 2e^{-M_1s}(1 + W(x)).
\]

By the triangle inequality, the fact that any \( \phi \in \mathcal{F} \) satisfies for some \( C > 0 \) that \( \phi \leq CW, \) together with (2.4.15) and Proposition 2.4.7, for all \( (s, x) \in [0, \infty) \times \mathbb{R}^d, \)

\[
\sup_{\phi \in \mathcal{F}} \left| \int_{\mathbb{R}^d} \phi(x') (p_{1+s}(x, d x') - q(d x')) \right| \leq \sup_{\phi \in \mathcal{F}} \left| \int_{\mathbb{R}^d} \phi(x') (p_{1+s}^1(x, d x') - p_{1+s}^1(x, d x')) \right| + \sup_{\phi \in \mathcal{F}} \left| \int_{\mathbb{R}^d} \phi(x') (p_{1+s}^1(x, d x') - q(d x')) \right| \leq \int_{\mathbb{R}^d} C(1 + W(x')) |p_{1+s}^1(x, d x') - p_{1+s}^1(x, d x')| + \sup_{\phi \in \mathcal{F}} \left| \int_{\mathbb{R}^d} \phi(x') (p_{1+s}^1(x, d x') - q(d x')) \right| \leq Ce^{-M_1s}(1 + W(x)) + Ce^{-cs} \int_{\mathbb{R}^d} (1 + W(x'' - x')) q(d x'').
\]

In other words, Assumption 2.2.2 is satisfied for some \( C, c > 0 \) with \( g(x) = W(x) + \int_{\mathbb{R}^d} W(x'' - x'') q(d x'') \) and \( G(s) = Ce^{-cs}. \)

Moreover, similarly to Lemma 2.3.7, under Assumptions 2.3.2 and 2.3.1 one asserts that Assumption 2.2.3 is satisfied with \( h(x) = C \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (W(x'') + W(x'' - x')) q(d x') e^{-c|x-x''|^2} d x'' \) since we assumed that \( W \) is integrable against the specific Gaussians in the bounds (2.3.3) and (2.3.4). Alternatively, under Assumptions 2.3.2 and 2.3.1(b), similarly to Lemma 2.3.8 we can prove that Assumption 2.2.3 is satisfied with \( h(x) = Ce^{v|x|^2} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} (1 + W(x'') + W(x'' - x')) q(d x') e^{-c|x-x''|^2 - v|x''|^2} d x'' \).

Moreover, Assumption 2.2.1 is proved under both Assumptions 2.3.1(a) and 2.3.1(b) by similar arguments to those in Lemmas 2.3.7 and 2.3.8 respectively.

Finally, since Assumptions 2.2.1, 2.2.2 and 2.2.3 are satisfied, Theorem 2.2.4 gives us the claim. \( \square \)
Chapter 3

Gradient estimates for linear functions of solutions to McKean–Vlasov SDEs

We produce explicit bounds for the gradient of the solution to the backwards Kolmogorov equation associated to a McKean–Vlasov SDE. This is done in two parts: first by applying the result in Chapter 2 where we achieve decaying estimates; and second by obtaining a characterization of the derivative of the transition density with respect to the initial condition, which allows us to use a priori estimates to obtain decaying or linearly growing in time bounds of part of the space derivative.

3.1 Introduction

The object of study of this chapter is the following real–valued, $d$–dimensional stochastic process $(X_t)_{t \geq 0}$ satisfying a so called McKean–Vlasov SDE. That is, for $P_2(\mathbb{R}^d)$ representing the set of probability measure over $\mathbb{R}^d$ with finite second moment, given coefficients $b : \mathbb{R}^d \times P_2(\mathbb{R}^d) \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \times P_2(\mathbb{R}^d) \to \mathbb{R}^d \times \mathbb{R}^d$ and initial datum $\xi$ with $L(\xi) \in P_2(\mathbb{R}^d)$ together with a Brownian motion process $(B_s)_{s \geq 0}$, we consider $X^\xi$ solution of the following SDE:

$$dX^\xi_t = b(X^\xi_t, \mathcal{L}(X^\xi_t)) \, ds + \sigma(X^\xi_t, \mathcal{L}(X^\xi_t)) \, dB_s, \quad \forall s \in (0, \infty); \quad X^\xi_0 = \xi. \quad (3.1.1)$$

Existence and uniqueness results have been proved for Lipschitz coefficients on $\mathbb{R}^d \times P_2(\mathbb{R}^d)$ (see [55]), for linearly growing coefficients (see [75]) or under monotonicity or Lyapunov conditions (see [39, 44, 76]).

Our goal is to provide explicit uniform bounds for the gradient of $V(s, x) := \mathbb{E}[\phi(X^x_s)]$, where $(X^x_s)_{s \geq 0}$ is the process satisfying (3.1.1). To that end we introduce an auxiliary non–McKean–Vlasov SDE (recall the last section of Chapter 2), assuming that (3.1.1) has a unique weak solution $(X^\xi_s)_{s \geq 0}$. Indeed, we define the functions $\beta(s, x; \xi) := b(x, \mathcal{L}(X^\xi_s))$ and $\Sigma(s, x; \xi) := \sigma(x, \mathcal{L}(X^\xi_s))$ by just plugging in the law of the solution. Then, from (3.1.1) we obtain the following related SDE for a fixed $\tau \geq 0$ and any $(s, y) \in [\tau, \infty) \times \mathbb{R}^d$:

$$dY_{\tau, y}^s = \beta(s, Y_{\tau, y}^s, \mathcal{L}(Y_{\tau, y}^s)) \, ds + \Sigma(s, Y_{\tau, y}^s, \mathcal{L}(Y_{\tau, y}^s)) \, dB_s; \quad Y_{\tau, y}^\tau = y. \quad (3.1.2)$$

Due to uniqueness, it turns out that for each $x \in \mathbb{R}^d$, $X^x_s = Y_{\tau, y}^0,_{x; x}$ a.e., and so in particular the functions $V(s, x) := \mathbb{E}[\phi(X^x_s)]$ and $V(s, y; x) := \mathbb{E}[\phi(Y_{\tau, y}^0,_{y; x})]$, defined for $\phi$ a measurable...
function, satisfy $V(s, x) = V(s, y; x)_{|y=x}$ for all $x \in \mathbb{R}^d$. Consequently, bounds on the derivatives of $V$ can be obtained by exploiting uniquely the dynamic of $Y$, which is that of a usual non-autonomous SDE, in the sense that $\partial_x Y(s, x) = \partial_y V(s, y; x)_{|y=x} + \partial_y V(s, y; x)_{|y=x}$. Based on this reformulation, we achieve explicit uniform bounds for the gradient of $V(s, x) = \mathbb{E}[\phi(X^0_s \circ x)]$ by splitting the task in two parts.

The first part is based on Chapter 2 in that, under certain regularity and monotonicity conditions for the coefficients, we ensure that we can apply Theorem 2.2.4 to the auxiliary non-autonomous process $Y$ to obtain decay in time for the gradient $(1, \infty) \times \mathbb{R}^d \ni (s, x) \mapsto \partial_y V(s, y; x)_{|y=x}$. More explicitly, convergence of the law of $Y$ towards invariant measure and a Bismut–type formula for the gradient of $V$ together with certain a priori uniform moment estimates, allow us to conclude the two conditions required in Theorem 2.2.4. This result is collected in Theorem 3.2.5 and unlike in Chapter 2, we do not use non–classical results from the literature.

The second part is obtained in Theorem 3.3.4 due to a reformulation of the gradient (in $x$) of the transition probability of $Y$ under monotonicity conditions and additional assumptions over the existence and growth of derivatives of the coefficients.

Regarding the motivation of such bounds, recall that dynamics like (3.1.1) arise naturally as limit of interacting particle systems (we denote their solution by $\{Z^i_{Nt}\}_{i=1,...,N}$, where $N$ is the number of particles in the system). It turns out that a particular limiting behaviour of this system is that as $N$ is allowed to go to infinity, in any finite subset the particles become asymptotically independent of each other and in fact they converge weakly to i.i.d copies of the original McKean–Vlasov process $X_s^\xi$. Such a phenomenon is known by the name of strong propagation of chaos and it has been proved to have a time–dependent convergence rate of order $\mathcal{O}(1/\sqrt{N})$ for different degrees of generality allowed for the test functions and coefficients (see [48, 55, 56, 62], [61, Ch.1]). Recently, it has also been proposed a sharp uniform–in–time propagation of chaos order in [59] for drifts whose dependence on the measure is linear.

Now, there is a version of this phenomenon known by the name of weak propagation of chaos, which deals with the statistical behaviour of the empirical distribution of the particle system. Its explicit bounds are concerned with estimates of expressions of the form $|\mathbb{E}[\phi(X_T)] - \mathbb{E}[\phi(Z^{1,N}_T)]|$ for some measurable test function $\phi : \mathbb{R}^d \mapsto \mathbb{R}$ and some fixed time $T > 0$. Recent work presented in [38, Theorem 2.17] asserts time–dependent order $\mathcal{O}(1/\sqrt{T})$ for the weak propagation of chaos, in the case of Lipschitz coefficients and bounded diffusion. Time discretization analysis has been carried out in [57, Theorem 2.8], where the authors obtain time–dependent order $\mathcal{O}(1/T + h)$ for the associated step–$h$ Euler scheme. These estimates hold for linear in measure coefficients and were generalised later in [51] to still linear but possibly discontinuous drifts. Unfortunately, uniform estimates in time are challenging and have only recently been proved on the torus in [60] for test functions on measure spaces, additive noise and drift belonging to some specific family such as the ones with small dependence on the measure. Their method relies on uniform and decaying in time estimates for the derivatives of the solution to the Master equation. Combining Theorems 3.2.5 and 3.3.4, we obtain in Corollary 3.4.1 decaying estimates in time of the gradient of $V$, i.e. the solution to the Master equation in the case that the test function is linear, on the whole $\mathbb{R}^d$ and without explicitly requiring linearity in the measure component for the coefficients. We must say that these are not sufficient for their weak error analysis since second order derivatives estimates are also needed. Further work is required for our methodology to be
used for obtaining decaying or uniform second order derivative estimates for \( V \).

However, as opposed to the calculus on measure spaces techniques that are used in \([38,60]\), for some cases it is enough to use regular PDE techniques when obtaining uniform estimates for the weak propagation of chaos (see Chapter 4 and \([57]\)). And these are based on the uniform decay of the order one and two space derivatives of the real solution (in variables \( s, y \)) to the auxiliary backward Kolmogorov equation, i.e.

\[
[0, \infty) \times \mathbb{R}^d \times \mathbb{R}^d \ni (s, y, x) \rightarrow V(s, y; x) = \mathbb{E}[\phi(y_s^0, y; x)],
\]

first of which can be obtained using only Theorem 3.2.5. In the case that the coefficients are continuous, bounded (in the measure component) and monotonic, we present in Example 3.1.5 an approach consisting in the combination of Theorem 3.1.4 and \([25, \text{Theorem 9.4.2, Remark below display (9.4.18)}]\) in order to obtain the uniform decay of these two space derivatives. In other words, our results allow the generalization of the available results for weak propagation of chaos to uniform estimates and not necessarily linear-in-measure coefficients.

### 3.1.1 Recall: Gradient in the first space component

This section is a reminder of Theorem 2.2.4, where the assumptions and different scenarios covered by the result are studied in more detail.

For \( y \in \mathbb{R}^d \) and \( \tau \geq 0 \), we consider the following non-autonomous SDE for given coefficients \( \beta : [\tau, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \), \( \Sigma : [\tau, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \times \mathbb{R}^d \):

\[
Y_{\tau+s}^y = y + \int_\tau^{\tau+s} \beta(t, Y_t^y) \, dt + \int_\tau^{\tau+s} \Sigma(t, Y_t^y) \, dB_t, \quad s \geq 0.
\]  

(3.1.3)

In order to avoid cumbersome notation, we will be using the following convention \( Y_{0+s}^y = Y_s^y \) for all \((s, y) \in [0, \infty) \times \mathbb{R}^d\).

To formulate the assumptions let us fix \( \mathcal{S} \subseteq B(\mathbb{R}^d) \) and \( N \in \mathbb{N} \).

First we require some regularity for the transition density of the studied process.

**Assumption 3.1.1**. For all \( x \in \mathbb{R}^d \), \( \tau \geq 0 \), the equation (3.1.3) has a solution \([\Omega, \mathbb{P}, (B_s)_{s \geq \tau}, (\mathcal{F}_s)_{s \geq \tau}, (Y_s^{x, T})_{s \geq \tau}]\) unique in the sense of probability. Moreover, we suppose that this process admits a density: \((s, x') \rightarrow p_{t+s}^{I}(x, x')\). Finally, if \( \mathcal{S} \not\subseteq C(\mathbb{R}^d) \), for any \( 0 \leq n \leq N \) there exists \( \delta > 0 \) such that

\[
\sup_{\phi \in \mathcal{S}} \int_{\mathbb{R}^d} |\phi(x'')| \partial_x^n p_0^I(x, x'')|^{1+\delta} \, dx'' < \infty, \quad \forall (s, x) \in (0, \infty) \times \mathbb{R}^d.
\]

Second, we assume that the transition density “remains close” to a static measure.

**Assumption 3.1.2** (Sticking to a measure). There exist \( q \in \mathcal{P}(\mathbb{R}^d) \), \( g : \mathbb{R}^d \rightarrow \mathbb{R}_+ \) and \( G : \mathbb{R}_+ \rightarrow \mathbb{R}_+ \) such that for all \((s, x) \in (0, \infty) \times \mathbb{R}^d\):

\[
\sup_{\phi \in \mathcal{S}} \left| \int_{\mathbb{R}^d} \phi(x') (p_{t+s}^{I}(x, dx') - q(dx')) \right| \leq g(x) G(s).
\]

This assumption can be obtained by means of tailoring the family of test functions in such a way that convergence to invariant measure in Wasserstein–\( m \) distance or Total Variation distance would be enough (in the case that \( \lim_{s \rightarrow \infty} G(s) = 0 \)) to obtain Assumption 3.1.2.
We consider here two alternatives for guaranteeing this type of convergence: Lyapunov conditions (see Appendix B.2) and monotonicity conditions (see Lemma 3.2.11). Since we focus on the latter in the reminder of the chapter, let us recall that (uniform and strict) monotonicity condition holds for the coefficients of (3.1.3) if there exists some $p \in \mathbb{N}$ and $M_1 > 0$ such that for all $t \in [0, \infty)$ and $x, x_0 \in \mathbb{R}^d$,

$$
(x - x_0, \beta(t, x) - \beta(t, x_0)) + \frac{p - 1}{2} |\Sigma(t, x) - \Sigma(t, x_0)|^2 \leq -M_1 |x - x_0|^2. \quad (3.1.4)
$$

Finally, we assume finite time regularity of the transition density, which we later translate into infinite time “good behaviour” for the space derivatives of the function $V$.

**Assumption 3.1.3** (Smoothness and integrability of the transition density w.r.t. the starting point). For any $0 \leq n \leq N$, we assume $(s, x, x',) \mapsto \partial^n_x p_t^0(x, x')$ exist and are continuous in $(x, x')$. Moreover, there exists $h : \mathbb{R}^d \to \mathbb{R}_+$ such that for $g$ satisfying Assumption 3.1.2 and any $1 \leq n \leq N$, the following is satisfied:

$$
\int_{\mathbb{R}^d} g(x'') |\partial^n_x p_t^0(x, x'')| dx'' \leq h(x), \quad \forall x \in \mathbb{R}^d.
$$

This assumption can be obtained as a consequence of bounds of the derivatives of the transition density with respect to the initial condition. This was studied by many from the PDE perspective (see e.g [25], [26]), the most common approach being the parametrix technique. In particular, [25, Theorem 9.4.2, Remark below display (9.4.18)] assume in addition to $C^2(\mathbb{R}^d) \ni u \mapsto Lu = \frac{1}{2} \text{tr}(\Sigma \Sigma^* \partial_x^2 u) + \beta \partial_x u$ being uniformly elliptic over $[0, \infty) \times \mathbb{R}^d$ (i.e there exists $\kappa > 0$ such that for all $(t, x, \xi) \in [0, \infty) \times \mathbb{R}^d \times \mathbb{R}^d$, we have $\xi^* \Sigma \Sigma^* (t, x) \xi \geq \kappa |\xi|^2$), $\beta, \Sigma \in C_b^{1,2}([0, 1] \times \mathbb{R}^d)$ and $\beta, \Sigma \in C_b^{1,2}([0, \infty) \times \mathbb{R}^d)$.

Due to the restrictive nature of the assumptions for these “off–the–shelf” results when combined with Lyapunov or monotonicity conditions, we present in the next section an alternative approach based uniquely on monotonicity and which is achieved due to a Bismut–Elworthy–Li type of formula. However, this approach results in estimates for just the gradient of $V$. Hence, despite being restrictive, for simpler scenarios one can combine the available PDE results with our approach and obtain estimates for higher order derivatives than the ones we obtain with the novel results we present on this Chapter (see Example 3.1.5).

**Theorem 3.1.4.** *Let $\phi \in \mathcal{S}$, $(Y_s^x)_{s \geq 0}$ be the unique (in law) solution of (3.1.3) and

$$
V(s, x) := \int_{\mathbb{R}^d} \phi(x') p_t^0(x, x') dx' = E[\phi(Y_s^x)].
$$

If Assumptions 3.1.1, 3.1.2 and 3.1.3 hold then, for all $1 \leq n \leq N$,

$$
|\partial^n_x V(s, x)| \leq h(x) G(s), \quad \text{for all } (s, x) \in (1, \infty) \times \mathbb{R}^d.
$$

Next we show an example which satisfies the assumptions of [26] and the monotonicity condition (3.1.4). In other words, Assumptions 3.1.1, 3.1.2 and 3.1.3 hold (for $N = 2$).

**Example 3.1.5.** Consider the Ornstein–Uhlenbeck process $(X_t)_{t \geq 0}$ given by the following SDE:

$$
dX_t^x = -X_t^x \alpha(E[X_t^x]) dt + \sqrt{2} \sigma dB_t; \quad X_0^x = x,
$$

58
where $\sigma > 0$ and $\alpha : \mathbb{R}^d \to \mathbb{R}$ a twice differentiable function such that there exists $M > M_0 > 0$ satisfying $M_0 \leq |\alpha(x)| \leq M, \ \forall x$. Additionally, for a fixed value $2 \leq m \in \mathbb{N}$, let us consider test functions $\phi : \mathbb{R}^d \to \mathbb{R}$ such that there exists $C > 0$ for which the following is satisfied:

$$|\phi(x) - \phi(y)| \leq C(1 + |x|^{m/2} + |y|^{m/2})|x - y|, \ \forall x, y \in \mathbb{R}^d.$$  \quad (3.1.6)

We next show that the process $(Y_t^{y,x})_{t \geq 0}$ satisfying

$$dY_t^{y,x} = -Y_t^{y,x} \alpha(\mathbb{E}[X_t^{y,x}]) dt + \sqrt{2\sigma} dB_t; \quad Y_0^{y,x} = y \in \mathbb{R}^d; \quad (3.1.7)$$

(associated process which has constant diffusion $\sqrt{2\sigma}$ and drift defined by $\beta(t, x) = -x\alpha(\mathbb{E}[X_t^x])$) and functions $\phi$ with the property stated above, satisfy all the assumptions for Theorem 3.1.4.

- On one hand, we have the monotonicity condition (3.1.4): for any $x, x_0 \in \mathbb{R}^d$ there exists $M_1 > 0$ such that for any $p \in \mathbb{N}$

$$\langle x - x_0, \beta(t, x) - \beta(t, x_0) \rangle + \frac{p-1}{2} |\Sigma(t, x) - \Sigma(t, x_0)|^2 = \langle x - x_0, -x\alpha(\mathbb{E}[X_t^x]) + x_0\alpha(\mathbb{E}[X_t^{x_0}]) \rangle$$

$$\leq -M_1 |x - x_0|^2.$$  

Next, notice that monotonicity assumptions in the shape of (3.1.4) imply exponential convergence to an invariant measure in any Wasserstein distance (see Lemma 3.2.10). Moreover, for test functions covered in this example and having convergence in Wasserstein distance to an invariant measure, we prove in Lemma 3.2.11 that Assumptions 3.1.2 is satisfied for some $C, c > 0$ with $s \mapsto G(s) = e^{-cs}$ and $x \mapsto g(x) = C(1 + |x|^m)$.

- On the other hand, since $\alpha$ is bounded and $\sigma > 0$ is constant, the regularity assumptions required by [25, Theorem 9.4.2, Remark below display (9.4.18)] hold. And, in Lemma 2.3.7 we proved that [25, Theorem 9.4.2, Remark below display (9.4.18)] imply that Assumption 3.1.1 and Assumption 3.1.3 hold for some $C, c > 0$ and

$$h(x) = \begin{cases} 
    Ce^{|x|^2}, & \text{if } p < m \\
    C(1 + |x|^{p/m}), & \text{if } p \geq m; \\
\end{cases} \quad \text{for all } x \in \mathbb{R}^d.$$

Hence, the conclusion of Theorem 3.1.4 holds. In particular, for fixed $x \in \mathbb{R}^d$, the first two derivatives in the $y$–space variable of $(1, \infty) \times \mathbb{R}^d \ni (t, y) \mapsto V(t, y; x) = \mathbb{E}[\phi(Y_t^{y,x})]$ decay exponentially for any $\phi$ satisfying (3.1.6) and for $(Y_t^{y,x})_{t \geq 0}$ satisfying (3.1.7).

Later in this section we justify again the construction of such an auxiliary process and we prove that as a consequence of the fact that $X_t^x = Y_t^{x,x}$, a.s (see Lemma C.2.1) for all $(t, x) \in (0, \infty) \times \mathbb{R}^d$ we have that $V(t, x; x) = \mathbb{E}[\phi(X_t^x)]$. Therefore, the bounds of the derivatives $\partial_y V(t, y; x)|_{y=x}$ and $\partial_y^2 V(t, y; x)|_{y=x}$ will carry over consequences for $\partial_x \mathbb{E}[\phi(X_t^x)]$ and $\partial_x^2 \mathbb{E}[\phi(X_t^x)]$.  

59
3.2 Derivative with respect to the initial data of the auxiliary backward Kolmogorov function

In this section we provide a tractable set of assumptions which allows us to apply Theorem 3.1.4 (because they imply Assumptions 3.1.1, 3.1.2 and 3.1.3) to the process $X$ defined by (3.1.1) without relying on the results from [25] and [26]. Our alternative is based on Bismut’s formula for the derivative, to which end we consider the following set of assumptions:

**Assumption 3.2.1.** [Regularity and growth coefficients]

$b, \sigma \in \mathcal{C}^{2,2}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d))$ and there exists $m \in \mathbb{N}$, $m \geq 2$ such that for fixed $x \in \mathbb{R}^d$,

(i) $\mathbb{R}^d \ni x' \mapsto \sup_{t \in (0,\infty)} b(x', \mathcal{L}(X_t^x)) \in B_m(\mathbb{R}^d)$ and $\mathbb{R}^d \ni x' \mapsto \sup_{t \in (0,\infty)} (\partial_t \sigma)(x', \mathcal{L}(X_t^x)) \in B_m(\mathbb{R}^d)$;

(ii) $\mathbb{R}^d \ni x' \mapsto \sup_{t \in (0,\infty)} \sigma(x', \mathcal{L}(X_t^x)) \in B_{m/2}(\mathbb{R}^d)$ and $\mathbb{R}^d \ni x' \mapsto \sup_{t \in (0,\infty)} (\partial_t \sigma)(x', \mathcal{L}(X_t^x)) \in C_b(\mathbb{R}^d)$.

Moreover, for all $(t, x') \in (0, \infty) \times \mathbb{R}^d$ the matrix inverse of $\sigma(x', \mathcal{L}(X_t^x))$ exists, is measurable and $\sigma^{-1}(x', \mathcal{L}(X_t^x)) \in B_{(m/2 - 1)}(\mathbb{R}^d)$.

**Assumption 3.2.2.** The differential operator $L$ defined by

$$C^2(\mathbb{R}^d) \ni u \mapsto Lu = L(\cdot, t; x)u = \frac{1}{2} \text{tr}(\sigma \sigma^*(\cdot, \mathcal{L}(X_t^x)) \partial_x^2 u) + b(\cdot, \mathcal{L}(X_t^x)) \partial_x u$$

is uniformly elliptic over $[0, \infty) \times \mathbb{R}^d$, i.e. there exists $\kappa > 0$ such that for all $(t, x, x', \xi) \in [0, \infty) \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d$, we have $\xi^* \sigma \sigma^*(x', \mathcal{L}(X_t^x)) \xi \geq \kappa |\xi|^2$.

**Assumption 3.2.3.** [Well posedness and existence of invariant measure] Given a fixed natural number $p \geq 4m + 1$ (where $m$ is the same as in Assumption 3.2.1 and the definition (3.2.1) of $\mathcal{S}_m$ below) and any $(x, \mu), (x_0, \mu_0) \in \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$, there exist some $M_1 > 2M_0 > 0$ such that for any coupling $\Pi$ between $\mu$ and $\mu_0$,

$$\langle -x_0, b(x, \mu) - b(x_0, \mu_0) \rangle + \frac{p - 1}{2} |\sigma(x, \mu) - \sigma(x_0, \mu_0)|^2 \leq -M_1 |x - x_0|^2 + M_0 \int_{\mathbb{R}^d} |x - x_0|^2 d\Pi(x, x_0).$$

**Definition 3.2.4.** Given a function $\phi : \mathbb{R}^d \to \mathbb{R}$ and a real number $\tilde{m} \geq 0$, we say that it satisfies the local Lipschitz condition of order $\tilde{m}$ if there exists $C > 0$ such that for all $x, y \in \mathbb{R}^d$,

$$|\phi(x) - \phi(y)| \leq C(1 + |x|^{\tilde{m}} + |y|^{\tilde{m}}) |x - y|.$$  

Finally, we restrict ourselves to $\phi$ within the family of functions satisfying the local Lipschitz condition of order $m/2$ ($m$ given by Assumption 3.2.1) in Definition 3.2.4, i.e. $\phi$ in

$$\mathcal{S}_m := \{ \phi : \mathbb{R}^d \to \mathbb{R} | \exists C > 0 \text{ s.t. } \forall x, y \in \mathbb{R}^d, |\phi(x) - \phi(y)| \leq C(1 + |x|^{m/2} + |y|^{m/2}) |x - y| \}. \quad (3.2.1)$$

We prove in Theorem 3.2.11 that for this family of functions, Assumption 3.1.2 with a decay in time (i.e. $\lim_{s \to 0} G(s) = 0$) reduces to the better studied convergence to invariant measure in 2–Wasserstein distance.

As we have seen in Example 3.1.5, in order to apply Theorem 3.1.4, we need to reframe our McKean–Vlasov SDE problem in terms of a non-autonomous SDE of the form (3.1.2). This is
done by taking advantage of the uniqueness of solution to (3.1.1) guaranteed by Assumptions 3.2.1 and 3.2.3. In other words: due to uniqueness and existence of solution, for any \( \xi \) square integrable, there exists a solution to (3.1.1): \( \mathcal{L}(X^\xi_t) \) (see [76, Theorem 6.1]). It can afterwards be plugged into the coefficients \( b \) and \( \sigma \) in order to obtain

\[
\beta(y, s; \xi) = b(y, \mathcal{L}(X^\xi_t)) \quad \text{and} \quad \Sigma(y, s; \xi) = \sigma(y, \mathcal{L}(X^\xi_t))
\]

respectively. Hence, from (3.1.1)—for any \( \tau \in (0, \infty) \), \( \xi \) square integrable random variable and \( y \in \mathbb{R}^d \)—we obtain the following related SDE on \([\tau, \infty)\):

\[
dY^\tau_{s, y; \xi} = \beta(s, Y^\tau_{s, y; \xi}; \xi) ds + \Sigma(s, Y^\tau_{s, y; \xi}; \xi) dB_s, \quad Y^\tau_{\tau, y; \xi} = y. \quad (3.2.2)
\]

We can now define, for any \( y \in \mathbb{R}^d \), the function \( \mathbb{E}[\phi(Y^0_{s, y; \xi})] \) to which we can apply Theorem 3.1.4 in order to obtain decay or at least uniform boundedness in time of \( (s, x) \mapsto \partial_y V(s, y; x)|_{y=x} \). It remains to check that the bounds we obtain remain valid in some sense for the function of interest \( V(s, x) = \mathbb{E}[\phi(X^x_{s, y; \xi})] \), where we are given a test function \( \phi \in \mathcal{F}_m \) defined in (3.2.1). This is consequence of Lemma C.2.1, where we prove that \( X^x_{s, y; \xi} = Y^0_{s, y; \xi} a.s \) for all \( s, x \), meaning that \( V(s, x) = V(s, y; x)|_{y=x} \) for all \( s, x \). However, notice that derivatives-wise and as a consequence of the Chain Rule, the full space derivative of \( V \) can be split as

\[
\partial_x V(s, x) = \partial_y V(s, y; x)|_{y=x} + \partial_x V(s, y; x)|_{y=x}, \quad (3.2.3)
\]

and Theorem 3.1.4 only gives bounds for \( (s, y) \mapsto \partial_y V(s, y; x) \) for any given \( x \in \mathbb{R}^d \). Similarly, for higher order derivatives we need the crossed derivatives as well. The good news is that for purposes such as weak error estimates, it is enough to obtain bounds for just \( \partial_y V(s, y; x)|_{y=x} \) (see Chapter 4 and [57]). More explicitly, after dismembering the expression into more approachable terms using Itô’s formula and the backward Kolmogorov PDE, some of them include a space derivative of \( V \) but only after the law of the studied process has been fixed. Meaning that we only care about bounding \( \partial_y V(s, y; x)|_{y=x} \) with an integrable time decay in order to obtain the claimed uniform in in time weak error estimates.

Next we state the main result of this Section.

**Theorem 3.2.5.** Let Assumptions 3.2.1,3.2.2 and 3.2.3 hold and consider \( Y(s)_{s \geq 0} \) the stochastic process solution to (3.2.2). Given \( \phi \in \mathcal{F}_m \) (defined in (3.2.1) for \( 2m + 5 \leq p \)) the function \( (0, \infty) \times \mathbb{R}^d \ni (s, y) \mapsto V(s, y; x) = \mathbb{E}[\phi(Y^x_{s, y; \xi})] \) satisfies the following for some \( C, c > 0 \):

\[
|\partial_y V(s, y; x)| \leq C(1 + |x|^{2m} + |y|^{2m})e^{-cs}, \quad \text{for all} \quad (s, x, y) \in (1, \infty) \times \mathbb{R}^d \times \mathbb{R}^d.
\]

**Proof.** In Theorem 3.2.9 we show how Assumptions 3.2.1,3.2.2 and 3.2.3 imply that Assumption 3.1.3 holds for \( N = 1 \) and any \( x \in \mathbb{R}^d \) fixed and with \( g(y) = C(1 + |x|^{m} + |y|^m), h(y) = C(1 + |x|^{2m} + |y|^{2m}) \) for some constant \( C > 0 \). In other words,

\[
\int_{\mathbb{R}^d} (1 + |x|^m + |y|^m)|\partial_y P^0_t(y, x'; x)| dx' \leq C(1 + |x|^{2m} + |y|^{2m}).
\]

And when repeating the same argument as in Theorem 3.2.9 for the integrand raised to a certain power \( 1 < 1 + \delta < 2 \), we conclude in the same way that for any \( \phi \in \mathcal{F}_m \) (and therefore
of at most \(m^{th}\)-order polynomially growing function
\[
\int_{\mathbb{R}^d} |\phi(x'')\partial_y p_t^0(y, x'', x)|^{1+\delta} dx'' < \infty.
\]

The difference is that in order to do so, we need uniform moment estimates up until order \((m + 2)(1 + \delta)\). We then conclude that Assumption 3.1.1 holds as well.

Moreover, in Theorem 3.2.11 we prove that with regular enough coefficients and Assumption 3.2.3, we have Assumption 3.1.2 holding with \(G(s) = e^{-cs}\).

Altogether, the necessary conditions for the result presented in Theorem 3.1.4 to hold are satisfied.

\[\square\]

### 3.2.1 Smoothness and integrability of derivative of the density w.r.t starting point

This section is devoted to obtaining Assumption 3.1.3 from Assumptions 3.2.1, 3.2.2 and 3.2.3.

In Chapter 2, we provided examples satisfying Assumption 3.1.3 by obtaining uniform bounds in finite time intervals for the derivatives of the transition density and, although enough, what we actually need is the existence of some function \(h\) such that
\[
\int_{\mathbb{R}^d} (1 + |x''|^{m})|\partial_x p_t^0(x, x'')|dx'' \leq h(x), \quad \forall x \in \mathbb{R}^d.
\]

One of the benefits of this observation is that although our object of study is the solution to a backward Kolmogorov equation (backward Kolmogorov function), in the literature most of the results are somehow presented for the forward Kolmogorov equation. And the fact that we are looking at the problem in the weak sense allows us to make use of the literature on both equations when proving the main theorem of this section, Theorem 3.2.7.

Let us discuss informally the idea behind this section. Fixed \(x \in \mathbb{R}^d\) and given \((Y_s^{T,y;x})_{s \in [\tau, 1]}\) solution to (3.2.2), for regular enough coefficients and regular enough \(\phi\), we consider the function \(V(s, y; x) = \mathbb{E}[\phi(Y_s^{T,y;x})]\). We next define the function \([0, (1 - \tau)] \times \mathbb{R}^d \ni (t, y) \mapsto v(t, y; x) := V(1 - t, y; x)\), which satisfies as a consequence of \(V\) satisfying the backward Kolmogorov equation (see [5]):

\[
\begin{align*}
\partial_t v(t, y; x) + L(t, y; x)v(t, y; x) &= 0, & (t, y) \in [0, 1 - \tau] \times \mathbb{R}^d \\
v(1 - \tau, y; x) &= \phi(y), & y \in \mathbb{R}^d; \\
\end{align*}
\]

where recall that \(L\) is the infinitesimal generator associated to the SDE, i.e. it is defined for any \(f \in C^2(\mathbb{R}^d)\) as
\[
(t, \infty) \times \mathbb{R}^d \ni (t, y) \mapsto L(t, y; x)f(y) = \frac{1}{2}\text{tr}\{\Sigma(t, y; x)\partial^2_y f(y)\Sigma^*(t, y; x)\} + \langle \beta(t, y; x)\partial_y f(y) \rangle.
\]

Next we use Itô’s formula on the process \((v(t, Y_t^{1, y;x}; x))_{t \in [0, 1 - \tau]}\). We then use the equation (3.2.4) to simplify the resulting expression and we substitute the terminal condition. Finally, for the particular terminal condition \(y \mapsto \phi(y) = \delta_{x''}(y)\) with \(x'' \in \mathbb{R}^d\) fixed (keep in mind that this is an informal discussion), and for the fixed parameter \(x \in \mathbb{R}^d\) we have for all
For the remainder of this section we will assume without loss of generality that $\tau = 0$.

Next we recall Bismut–Elworthy–Li formula (see [77] and Theorem 3.2.7), which claims that under enough assumptions

$$\partial_y v(t, y; x) = \mathbb{E}\left[ \frac{\phi(Y_t^{y,x})}{1-t} \int_t^1 \Sigma^{-1}(s, Y_s^{t,y,x}; x) Z_s^{t,y,x} dB_s \right],$$

where $Z_s^{t,y,x} := \partial_y Y_s^{t,y,x}$ is the derivative process, i.e. the well defined process (see [78]) solution to the following associated SDE:

$$dZ_s^{t,y,x} = (\partial_y \beta)(s, Y_s^{t,y,x}) Z_s^{t,y,x} ds + (\partial_y \Sigma)(s, Y_s^{t,y,x}; x) Z_s^{t,y,x} dB_s, \quad Z_0^{t,y,x} = I_d. \quad (3.2.6)$$

Applying Bismut–Elworthy–Li formula to $v$ and translating its implication to $p$ with (3.2.5), we conclude that the derivative of $p$ has the following representation

$$\partial_y p_1^0(y, x''; x) = \mathbb{E}\left[ \frac{\delta_{X'}^y(Y_t^{t,y,x})}{1-t} \int_t^1 \Sigma^{-1}(s, Y_s^{t,y,x}; x) Z_s^{t,y,x} dB_s \right], \quad (3.2.7)$$

which, together with uniform moment bounds for all processes involved, give hope for bounding the derivative.

Clearly, because of the Dirac delta, the expression (3.2.7) doesn’t have any meaning in the usual sense. However, we will consider it in the weak sense. Indeed, in the next lemma we show convergence to $p_1^0(y, ; x)$ of the solution to the version of the backwards PDE (3.2.4) where we smooth the terminal condition:

$$\begin{cases}
\partial_t p_{1,\epsilon}^0(y, x''; x) + L(t, y; x) p_{1,\epsilon}^0(y, x''; x) = 0, & (t, y) \in [0,1) \times \mathbb{R}^d \\
p_{1,\epsilon}^0(y, x''; x) = \zeta''_\epsilon(y), & y \in \mathbb{R}^d; 
\end{cases} \quad (3.2.8)$$

where $\zeta''_\epsilon(\cdot)$ is a mollified version of $\delta_{X'}(\cdot)$ (for more details see proof of Lemma 3.2.6).

**Lemma 3.2.6.** Given $m \geq 2$, let Assumptions 3.2.1, 3.2.2 and 3.2.3 hold with $p \geq 2m + 1$ and $\psi$ be a function satisfying Definition 3.2.4 with order $m/2$. For any fixed $x, y \in \mathbb{R}^d$ let $z \mapsto p_1^0(y, z; x)$ be the density of $Y_1^{t,x}$, solution at time 1 to (3.2.2) and, once fixed $x, z \in \mathbb{R}^d$, for every $\epsilon \in (0,1)$ let $y \mapsto p_{1,\epsilon}^0(y, z; x)$ be the solution to the PDE (3.2.8). Then,

$$\lim_{\epsilon \to 0} \int_{\mathbb{R}^d} |\psi(x'')| |\partial_y p_1^0(y, x''; x) - \partial_y p_{1,\epsilon}^0(y, x''; x)| dx'' = 0, \quad \forall (y, x) \in \mathbb{R}^d \times \mathbb{R}^d. \quad (3.2.9)$$

**Proof.** Take the standard family of mollifiers defined for $\epsilon \in (0,1)$ by:

$$\rho_\epsilon(x) = \begin{cases}
\frac{C}{\epsilon^2} \exp\left(\frac{x^2}{|x|^2 - \epsilon^2}\right), & \text{if } |x| < \epsilon \\
0, & \text{if } |x| \geq \epsilon,
\end{cases}$$

63
where $C > 0$ is the normalization constant which makes them integrate to 1. This is a family of positive, symmetric mollifiers with bounded derivatives and compact support (denoted by $\text{supp}(\rho_c)$) of length $2\epsilon$.

Consider now the solution of the backward Kolmogorov PDE with smooth terminal condition result of the convolution of $\delta_{x''}$ with $\rho_c$. Namely, for fixed $x'' \in \mathbb{R}^d$ and $\epsilon \in (0, 1)$, let us define $y \mapsto \zeta_{\epsilon}^{x''}(y) := (\delta_{x''} * \rho_c)(y)$, where $*$ stands for convolution operator and consider (3.2.8).

First, under Assumption 3.2.1, for any fixed $x, x'' \in \mathbb{R}^d$ and $\epsilon \in (0, 1)$, $(t, y) \mapsto E[\zeta_{\epsilon}^{x''}(Y_t^1, y, x)]$ belongs to $C^{1,2}([0, 1] \times \mathbb{R}^d)$ [79, Section 2.9, Theorem 10]. Then, we can apply Itô’s formula to $(\zeta_{\epsilon}^{x''}(Y_t^1, y, x))_{t \in [\tau, 1)}$ and obtain that $p_{1, \epsilon}^t(y, x''; x) = E[\zeta_{\epsilon}^{x''}(Y_t^1, y, x)]$ is a classical solution to (3.2.8).

Next we prove in two steps that (3.2.9) holds for any $\psi$ satisfying the local Lipschitz condition of order $m/2$ (see Definition 3.2.4).

**Step 1:** We claim that for all $t, y, x$,

$$
\int_{\mathbb{R}^d} \left| \psi(x'') \left( p_{1, \epsilon}^t(y, x''; x) - p_{1}^t(y, x''; x) \right) \right| dx'' \to 0, \text{ as } \epsilon \to 0.
$$

(3.2.10)

First, we notice that from the uniqueness of solution to the Fokker–Planck equation (second order parabolic PDE), for $y \in \mathbb{R}^d$ fixed, $x'' \mapsto p_{1, \epsilon}^t(y, x''; x)$ is also the density of the law of the random variable $Y_t^1, X_{\epsilon}^{x''}$, with $X_{\epsilon}^{x''} \sim \zeta_{\epsilon}^{x''}(z)dz$. Hence, under Assumption 3.2.3, Lemma 3.2.10 implies that for all $t \in [0, 1)$,

$$
\mathbb{E}[|Y_t^1, X_{\epsilon}^{x''} - Y_t^1, y, x|^m] \leq e^{-mM_1(1-t)} \mathbb{E}[|X_{\epsilon}^{y} - y|^m] = e^{-mM_1(1-t)} \int_{\mathbb{R}^d} |z - y|^m \zeta_{\epsilon}^{y}(z)dz

\leq e^{-mM_1(1-t)} \int_{\mathbb{R}^d} (2\epsilon)^m \zeta_{\epsilon}^{y}(z)dz = e^{-M_1m(1-t)} (2\epsilon)^m.
$$

(3.2.11)

Notice then that as $\epsilon \to 0$, we obtain $\mathbb{E}[|Y_t^1, X_{\epsilon}^{x''} - Y_t^1, y, x|^m] \to 0$. Similarly, but using Lemma B.1.3 instead, one obtains that

$$
\sup_{\epsilon \in (0, 1)} \mathbb{E}[|Y_t^1, X_{\epsilon}^{x''} - y|^m] \leq \sup_{\epsilon \in (0, 1)} C(1 + (e + |y|)^m) \leq C.
$$

And since for any $\psi$ satisfying the local Lipschitz condition of order $m/2$

$$
\int_{\mathbb{R}^d} \psi(x'') \left( p_{1, \epsilon}^t(y, x''; x) - p_{1}^t(y, x''; x) \right) dx''

\leq \mathbb{E}\left[ \psi(Y_t^1, X_{\epsilon}^{y}; x) - \psi(Y_t^1, y, x) \right]

\leq C \mathbb{E}\left[ \left( 1 + |Y_t^1, X_{\epsilon}^{y}; x|^m + |Y_t^1, y, x|^m \right) \right]^{1/2} \left[ \mathbb{E}\left[ |Y_t^1, X_{\epsilon}^{y}; x - Y_t^1, y, x|^2 \right] \right]^{1/2},
$$

we conclude that $\int_{\mathbb{R}^d} \psi(x'') \left( p_{1, \epsilon}^t(y, x''; x) - p_{1}^t(y, x''; x) \right) dx'' \to 0$, as $\epsilon \to 0$. However, a part from lacking a derivative, this is not the convergence in $L^1$ we claimed so we need an extra step: we apply Vitali’s Theorem (see [17, Chapter 4]), given that we have uniform (in $\epsilon$) integrability (in $x''$) of $|\psi(x'')|p_{1, \epsilon}^t(y, x''; x)$, to conclude (3.2.10).

So let us check that we have the necessary uniform integrability, i.e. the uniform integrability of the family $\{\psi(Y_t^1, X_{\epsilon}^{y}; x)\}_{\epsilon \in (0, 1)}$. We do so using De La Vallée Poussin Theorem
we bound the right hand side uniformly in the relevant parameters and the estimates are\(\epsilon\) together with Lemma 3.2.10 and uniform moment estimates imply: for all\(m\) and

\[\sup_{\epsilon \in (0,1)} \text{Sup} [\epsilon (Y_{1}^{t,x,h})] [1 + \delta] < \infty.\]

Indeed, notice that fixed\(\delta > 0\), first using the \(m^{th}\) order growth of any \(m/2\)–locally Lipschitz function, afterwards the moment bounds obtained in Lemma B.1.3 and finally the compact support of \((\rho_{\epsilon})\), we conclude:

\[\sup_{\epsilon \in (0,1)} \text{Sup} E(|\psi(Y_{1}^{t,x,h})|^{1 + \delta})\]

\[\leq C \left( 1 + \sup_{\epsilon \in (0,1)} \text{Sup} E\left[ Y_{1}^{t,x,h} |^{m(1 + \delta)} \right] \right) \leq C \left( 1 + |x|^{m(1 + \delta)} + \sup_{\epsilon \in (0,1)} E|\chi_{\epsilon}|^{m(1 + \delta)} \right)\]

\[\leq C \left( 1 + |x|^{m(1 + \delta)} + \sup_{\epsilon \in (0,1)} \int \int \int \int (\xi - \eta) + \int \int \int \int (\xi - \eta)^{m(1 + \delta)} \epsilon^{Y}_{\epsilon}(z) d\xi d\eta dq d\nu \right) \leq C (1 + |x|^{m(1 + \delta)} + |y|^{m(1 + \delta)}).

**Step 2:** In this step we prove that we can exchange limit in \(\epsilon\) and integral in \(x''\) with the derivative in \(y\), in order to obtain from (3.2.10):

\[\lim_{\epsilon \to 0} \int_{\mathbb{R}^{d}} |\psi(x'')| |\partial_{y} p_{1}^{0}(y, x'', x) - \partial_{y} p_{1}^{0}(y, x'', x)| d\xi d\eta\]

\[= \partial_{y} \lim_{\epsilon \to 0} \int_{\mathbb{R}^{d}} |\psi(x'')| |p_{1}^{0}(y, x'', x) - p_{1}^{0}(y, x'', x)| d\xi = 0.\]

We make a quick note of convenience. Since by the inverted triangle inequality we have

\[\frac{1}{h} \int_{\mathbb{R}^{d}} |\psi(x'')| |(p_{1}^{0}(y + he_{i}, x'', x) - p_{1}^{0}(y + he_{i}, x'', x)) - (p_{1}^{0}(y, x'', x) - p_{1}^{0}(y, x'', x))| d\xi d\eta\]

\[\leq \frac{1}{h} \int_{\mathbb{R}^{d}} |\psi(x'')| |(p_{1}^{0}(y + he_{i}, x'', x) - p_{1}^{0}(y + he_{i}, x'', x)) - (p_{1}^{0}(y, x'', x) - p_{1}^{0}(y, x'', x))| d\xi d\eta,

we bound the right hand side uniformly in the relevant parameters and the estimates are carried over to the left hand side.

First, we show that for \(h \in (0, 1), \{e_{i}\}_{i=1,...,d}\) canonical basis on \(\mathbb{R}^{d}\) and for fixed \(x, y \in \mathbb{R}^{d}\) and \(\psi\) satisfying the local Lipschitz condition of order \(m/2\), Cauchy–Schwarz inequality together with Lemma 3.2.10 and uniform moment estimates imply: for all \(i \in \{1,...,d\},

\[\sup_{\epsilon \in (0,1)} \frac{1}{h} \int_{\mathbb{R}^{d}} |\psi(x'')| \left( p_{1}^{0}(y + he_{i}, x'', x) - p_{1}^{0}(y + he_{i}, x'', x) \right) - \left( p_{1}^{0}(y, x'', x) - p_{1}^{0}(y, x'', x) \right) \right| d\xi d\eta\]

\[\leq \sup_{\epsilon \in (0,1)} \frac{1}{h} \text{Sup} E\left[ (Y_{1}^{t,x,h} |^{m} + Y_{1}^{t,x,h} ; |^{m} |^{2} )^{1/2} (Y_{1}^{t,x,h} \in Y_{1}^{t,x,h} ; |^{2} )^{1/2} \right] + \sup_{\epsilon \in (0,1)} \frac{1}{h} \text{Sup} E\left[ (Y_{1}^{t,x,h} + Y_{1}^{t,x,h} ; |^{2} )^{1/2} (Y_{1}^{t,x,h} - Y_{1}^{t,x,h} ; |^{2} )^{1/2} \right] + \sup_{\epsilon \in (0,1)} \frac{1}{h} \text{Sup} E\left[ (Y_{1}^{t,x,h} + Y_{1}^{t,x,h} ; |^{2} )^{1/2} (Y_{1}^{t,x,h} + Y_{1}^{t,x,h} ; |^{2} )^{1/2} \right] < \infty.

(3.2.13)
We consider first the limits in $\epsilon$ and $h$: by the Moore–Osgood Theorem on interchanging of limits (see [21, Theorem 1]), given the uniform bound (3.2.13) in $\epsilon$ and the pointwise convergence (fixed $h$) of

$$\frac{1}{h} \int_{\mathbb{R}^d} |\psi(x'')(|p_{1,c}^0(y + he_i, x'', x) - p_{1}^0(y, x'', x)|)dx'' \to$$

$$\partial_{x'} \int_{\mathbb{R}^d} |\psi(x'')(|p_{1,c}^0(y, x'', x) - p_{1}^0(y, x'', x)|)dx'',$n

as $h \to 0$, we exchange limits in $\epsilon$ and $h$, i.e. limit in $\epsilon$ and derivative in $y$ in (3.2.12).

We prove now the second limit exchange in between the derivative in $y$ and the integral in $x''$. Let us fix then $\epsilon > 0$ and $x \in \mathbb{R}^d$. Again, once checked uniform integrability, from Vitali’s Convergence Theorem we would get $L^1$ convergence of $\frac{1}{h} |\psi(x'')(|p_{1,c}^0(y + he_i, x'', x) - p_{1}^0(y, x'', x)|)dx'' \to |\psi(x'')| \partial_{y'} \int_{\mathbb{R}^d} |\psi(x'')(|p_{1,c}^0(y, x'', x) - p_{1}^0(y, x'', x)|)dx''$. So let us check uniform integrability, for which we fix $A \subseteq \mathbb{R}^d$ a Borel measurable set and consider the following:

$$\sup_{h \in (0,1)} \frac{1}{h} \int_A \psi(x'')(p_{1,c}^0(y + he_i, x'', x) - p_{1}^0(y, x'', x)) dx'' \leq \sup_{h \in (0,1)} \frac{1}{h} \int_A \psi(x'')((p_{1,c}^0(y + he_i, x'', x) - p_{1}^0(y, x'', x)) dx'' \leq \sup_{h \in (0,1)} \frac{1}{h} \int_A \psi(x'')((p_{1,c}^0(y + he_i, x'', x) - p_{1}^0(y, x'', x)) dx''). \tag{3.2.14}$$

(3.2.15)

Focusing first on the mollified solution and recalling the $m^{th}$ order growth of the test functions, the expression $p_{1,c}^0(y, x'', x) = \mathbb{E}[\xi_c(x') (Y_{1,1}^{y,x})]$ due to smoothness of $\xi_c$ and PDE (3.2.4) and finally the properties of $(\rho_c)$,

$$\sup_{h \in (0,1)} \frac{1}{h} \int_A |\psi(x'')(p_{1,c}^0(y + he_i, x'', x) - p_{1}^0(y, x'', x)) dx''$$

$$= \sup_{h \in (0,1)} \frac{1}{h} \int_A \psi(x'')\mathbb{E}[\xi_c(x') (Y_{1,1}^{y,x} + he_i;x)] - \mathbb{E}[\xi_c(x') (Y_{1,1}^{y,x})] |dx''$$

$$= \sup_{h \in (0,1)} \int_A \psi(x'') \int_{\mathbb{R}^d} \xi_c(x') \int_{\mathbb{R}^d} \xi_c(x') (z - r) \rho_c(r) dr \int_{\mathbb{R}^d} \psi(x') \int_{\mathbb{R}^d} \rho_c(r) (z - x') (z - r) \rho_c(r) dr |dx''$$

$$\leq C \epsilon \sup_{h \in (0,1)} \int_{\mathbb{R}^d} (1 + |x''|^{m}) |p_{1,c}^0(y + he_i, z + x'', x) - p_{1}^0(y, z + x'', x)) dx''$$

$$\leq C \epsilon \sup_{h \in (0,1)} \int_{\mathbb{R}^d} (1 + |x''|^{m}) |p_{1,c}^0(y + he_i, x'', x) - p_{1}^0(y, x'', x)) dx''$$

$$\leq C \epsilon \sup_{h \in (0,1)} \int_{\mathbb{R}^d} (1 + |x''|^{m}) |p_{1,c}^0(y + he_i, x'', x) - p_{1}^0(y, x'', x)) dx''$$
\[ \leq C e \sup_{h \in (0,1)} \frac{1}{h} \int_A (1 + |x''|^m) \left| p^0_1(y + he_i, x''; x) - p^0_1(y, x''; x) \right| dx''. \]

We therefore bounded (3.2.14) by the second term (3.2.15) (once we acknowledge the fact that \( \psi \in B_m(\mathbb{R}^d) \)). Moreover, similarly to (3.2.13),

\[
\sup_{h \in (0,1)} \frac{1}{h} \int_A (1 + |x''|^m) \left| p^0_1(y + he_i, x''; x) - p^0_1(y, x''; x) \right| dx'' \\
\leq \sup_{h \in (0,1)} \left( E \left[ 1_A \left( 1 + |Y^t_{1, y + he_i; x} + Y^t_{1, y; x}|^m \right) \right] \right)^{1/2}
\]

and therefore goes to 0 as the Lebesgue measure of the set \( A \) goes to 0. We then conclude uniform (in \( h \)) integrability (in \( x'' \)) of

\[
\left( \frac{1}{h} \left| \psi(x'') \right| \left( p^0_{1, c}(y + he_i, x''; x) - p^0_1(y + he_i, x''; x) - p^0_{1, c}(y, x''; x) + p^0_1(y, x''; x) \right) \right)_{h},
\]

and Vitali’s theorem can be used to exchange derivative in \( y \) and integral in \( x'' \).

Although possibly available in the literature, the author could not find a precise reference for a Bismut formula not using the boundedness of the drift and diffusion coefficients. One of the reasons might be the fact that, in absence of such boundedness, one must prove square integrability of the process \( \{ \partial_y \nu(s, Y^0_{r, y; x}, x) \Sigma(s, Y^0_{r, y; x}) \}_{r \in [0,1]} \). Based on Assumption 3.2.1 and for test functions \( \phi \in C^1_b(\mathbb{R}^d) \), this is exactly what we prove in Theorem 3.2.7 using a probabilistic approach.

**Theorem 3.2.7.** Let Assumptions 3.2.1, 3.2.2 and 3.2.3 hold and assume moreover that \( \phi \in C^1_b(\mathbb{R}^d) \). Let \( (Y^0_{r, y; x})_{r \geq 0} \) and \( (Z^0_{r, y; x})_{r \geq 0} \) be the solutions to (3.2.2) and (3.2.6) respectively. Then, for any \( x \in \mathbb{R}^d \) fixed, there exists a unique solution \( \nu(\cdot, y; x) \in C^{1,2}([0,1] \times \mathbb{R}^d) \) to (3.2.4) and it satisfies the following Bismut–Elworthy–Li formula:

\[
\partial_y \nu(1, y; x) = E \left[ \phi(Y^0_{1, y; x}) \left( \int_0^1 (\Sigma^{-1}(r, Y^0_{r, y; x}) T^0_{r, y; x} d B_r) \right)^* \right], \quad \forall y \in \mathbb{R}^d.
\]

**Proof.** First notice that, for being solution to the PDE (3.2.4), \( \nu(\cdot, y; x) \) belongs to \( C^{1,2}([0,1] \times \mathbb{R}^d) \) (see [79, Section 2.9, Theorem 10]), which is valid under Assumptions 3.2.1 and 3.2.2. Moreover, from (3.2.4) we know that for any \( r \geq 0 \) we can express \( \nu(r, Y^0_{r, y; x}; x) \) as follows:

\[
\nu(r, Y^0_{r, y; x}; x) = \nu(0, y; x) + \int_0^r \partial_y \nu(s, Y^0_{s, y; x}; x) \Sigma(s, Y^0_{s, y; x}) d B_s.
\]

Next, we prove the square integrability of \( (\partial_y \nu(s, Y^0_{s, y; x}; x) \Sigma(s, Y^0_{s, y; x}) \}_{r \geq 0} \). Notice that since \( \Sigma \) is assumed to have polynomial growth and we obtained uniform moments estimates in Lemma B.1.3, it is enough to prove that for \( x \) fixed, \( [0,1] \times \mathbb{R}^d \ni (s, y) \mapsto \partial_y \nu(s, y; x) \) is bounded. And the later, we will see next, is a consequence of being able to express \( \nu \) as \( \nu(s, y; x) = E[\phi(Y^0_{1, y; x})] \) (see (3.2.5)).

Similarly to Lemma 3.2.6, we notice that, due Lemma 3.2.10, for any \( h \in (0,1) \) and
\{e_i\}_{i=1,...,d} canonical base in \(\mathbb{R}^d\),

\[
\phi(z) \frac{1}{h}(p_i^1(y + he_i, z; x) - p_i^1(y, z; x))
\]

is uniformly (in \(h\)) integrable (in \(z\)). This means that the necessary conditions in Vitali’s Theorem are satisfied (see [17, Chapter 4]) and we obtain the following \(L^1\)-convergence

\[
\int_{\mathbb{R}^d} \phi(z) \frac{1}{h}(p_i^1(y + he_i, z; x) - p_i^1(y, z; x)) dz \to \int_{\mathbb{R}^d} \phi(z) \partial_y p_i^1(y, z; x) dz, \quad \text{as } h \to 0.
\]

In other words, \(\partial_y v(s, y; x) = \partial_y \mathbb{E}[\phi(Y_1^x)] = \mathbb{E}[\partial_y(\phi(Y_1^x))]\). Now, this together with the Chain Rule and Cauchy-Schwarz inequality lead to:

\[
|\partial_y v(s, y; x)| = |\mathbb{E}[\partial_y(\phi)(Y_1^x)]| \leq \sqrt{\mathbb{E}[|\partial_y(\phi)(Y_1^x)|^2]} \leq \sqrt{\mathbb{E}[|Z_1^x|^2]}^{1/2}.
\]

Next, the assumed boundedness of \(y \mapsto \partial_y \phi(y)\) and Lemmas B.1.3, B.1.6 allow us to conclude that there exists \(C > 0\) such that \(|\partial_y v(s, y; x)| \leq C\). All together implies that when applying Itô’s isometry we obtain the claimed square integrability:

\[
\mathbb{E}\left[\left(\int_0^r \partial_y v(s, Y_s^x; x) \sum(s, Y_s^x) dB_s\right)^2\right] \leq C \mathbb{E}\left[\int_0^r \left|\partial_y v(s, Y_s^x; x) \sum(s, Y_s^x)\right|^2 ds\right] < \infty.
\]

Hence, when taking expectations in (3.2.16), we obtain that for all \(r \geq 0\), \(\mathbb{E}[v(r, Y_r^0; x)] = v(0, y; x)\). Now, this together with another application of the Chain Rule, Itô’s isometry and the fact that \(v(r, y; x)\) is solution of (3.2.4) implies that:

\[
\partial_y v(0, y; x) = \mathbb{E}\left[\int_0^1 (\partial_y v(r, Y_r^0; x) Z_r) dr\right]
\]

\[
= \mathbb{E}\left[\left(\int_0^1 (\partial_y v(r, Y_r^0; x) \sum(r, Y_r^0; x) dB_r)\right)\left(\int_0^1 (\Sigma^{-1}(r, Y_r^0; x) Z_r^{0, y; x} Z_r^{0, y; x})* dB_r\right)\right]
\]

Moreover, since \(\Sigma^{-1}\) is has polynomial growth of at most order \(m/2 - 1\) and \(Z_r\) is square integrable by Lemma B.1.6, we obtain

\[
\mathbb{E}\left[\left(\int_0^1 (\Sigma^{-1}(r, Y_r^0; x) Z_r^{0, y; x} Z_r^{0, y; x})* dB_r\right)\right] = \mathbb{E}\left[\left(\int_0^1 (\Sigma^{-1}(r, Y_r^0; x) Z_r^{0, y; x} Z_r^{0, y; x})* dB_r\right)\right] = 0.
\]

And in fact,

\[
\partial_y v(0, y; x) = \mathbb{E}\left[v(1, Y_1^0; x)\left(\int_0^1 (\Sigma^{-1}(r, Y_r^0; x) Z_r^{0, y; x} Z_r^{0, y; x})* dB_r\right)\right]
\]

68
Theorem 3.2.9.
Given m

Corollary 3.2.8. Let Assumptions 3.2.1, 3.2.2 and 3.2.3 hold and 
Proof. For a fixed
In particular , the process
Definition 3.2.4 with order m
holds for
\[ \int R \psi \left( x, t \right) \left| \frac{\partial}{\partial x} \psi \left( x, t \right) \right| \ dx = \int R \psi \left( x, t \right) \left( \int_0^1 \left( \Sigma^{-1} \left( r, Y_r^{0, y, x}; x \right) Z_r^{0, y, x} \right) \ast dB_r \right) \ast.
\]
Notice that, due to the regularizing properties of the mollifiers, one can take for any \( \epsilon \in (0, 1) \) fixed, \( \phi = \zeta^{0}_\epsilon \). Then, fixed \( \epsilon \in (0, 1) \) and \( x'' \), \( x \in \mathbb{R}^d \), the conclusion to Theorem 3.2.7 holds for \( p_{1, \epsilon}(\cdot, x''; x) \), solution to (3.2.8).

Corollary 3.2.8. Let Assumptions 3.2.1, 3.2.2 and 3.2.3 hold and \( \psi \) be a function satisfying Definition 3.2.4 with order \( m/2 \) (\( m \leq (p-1)/2 \)). Fixed \( x, y \in \mathbb{R}^d \) and \( t \in (0, 1) \), let \( p^1_t(y, \cdot; x) \) be the density of \( Y^{1, y, x}_t \), solution at time 1 to (3.2.2), and \( (Z^1_s)_{s \geq t} \), solution to (3.2.6). Then,
\[ \int \partial_y p^0_t(y, z; x) \psi(z) \ dz = \mathbb{E} \left[ \psi(Y^0_1, y, x) \left( \int_0^1 \Sigma^{-1} \left( s, Y^0_s, x \right) Z^0_s \right) \ast dB_s \right]. \]  
(3.2.17)
Proof. First, based on the Bismut–Elworthy–Li Formula proved in Theorem 3.2.7 and Fubini’s Theorem:
\[ \int R \psi(x'') \partial_y p^0_{1, \epsilon}(y, x''; x) \ dx'' = \mathbb{E} \left[ \psi(Y^0_1, y, x'') \left( \int_0^1 \Sigma^{-1} \left( r, Y^0_r, x \right) Z^0_r \right) \ast dB_r \right]. \]  
(3.2.18)
Moreover, by Lemma 3.2.6, (3.2.18) can be used in the following way:
\[ \int R \psi(x'') \partial_y p^0_{1, \epsilon}(y, x''; x) \ dx'' = \lim_{\epsilon \to 0} \int R \psi(x'') \partial_y p^0_{1, \epsilon}(y, x''; x) \ dx'' \]
\[ = \lim_{\epsilon \to 0} \int R \psi(x'') \zeta^{0}_\epsilon \left( Y^0_1, y, x'' \right) dx'' \left( \int_0^1 \Sigma^{-1} \left( r, Y^0_r, x \right) Z^0_r \right) \ast dB_r \ast \]
\[ = \mathbb{E} \left[ \lim_{\epsilon \to 0} \int R \psi(x'') \zeta^{0}_\epsilon \left( Y^0_1, y, x'' \right) dx'' \left( \int_0^1 \Sigma^{-1} \left( r, Y^0_r, x \right) Z^0_r \right) \ast dB_r \ast \right]. \]
\[ = \mathbb{E} \left[ \psi(Y^0_1, y, x'') \left( \int_0^1 \Sigma^{-1} \left( r, Y^0_r, x \right) Z^0_r \right) \ast dB_r \ast \right]. \]
\[ \square \]

Theorem 3.2.9. Given \( m \geq 2 \), let Assumptions 3.2.1, 3.2.2 and 3.2.3 hold with \( p \geq 2m + 5 \) for the process \( (X^3_t)_{t \geq 0} \) solution to (3.1.1). Consider the density \( x'' \rightarrow p^0_t(y, x''; x) \) of the random variable \( Y^{0, y, x}_t \) given by (3.2.2). For any \( \psi \in \mathcal{F}_{2m} \) (see Definition 3.2.1) there exists \( C > 0 \) such that
\[ \int R \left| \psi(x'') \right| \left| \partial_y p^0_t(y, x''; x) \right| dx'' \leq C(1 + |y|^{2m} + |x|^{2m}). \]
In particular,
\[ \int R (1 + |x''|^{m}) \left| \partial_y p^0_t(y, x''; x) \right|_{y=x} \ dx'' \leq C(1 + |x|^{2m}), \ \forall x \in \mathbb{R}^d. \]
Proof. For a fixed \( \epsilon \in (0, 1) \), the algebraic inequality which for any \( a, b \in \mathbb{R} \) claims that \( |a| - |b| \leq |a - b| \) implies that for any \( \psi \) satisfying the local Lipschitz condition of order \( 2(m - 1) \)
(see Definition 3.2.4) and \(i = 1, \ldots, d\):

\[
\lim_{\epsilon \to 0} \int_{\mathbb{R}^d} |\psi(x'')| \|\partial_y p^0_1(y, x''; x)\| dx'' - \int_{\mathbb{R}^d} |\psi(x'')| \|\partial_y p^0_{1,\epsilon}(y, x''; x)\| dx'' \leq \lim_{\epsilon \to 0} \int_{\mathbb{R}^d} |\psi(x'')| \|\partial_y p^0_1(y, x''; x) - \partial_y p^0_{1,\epsilon}(y, x''; x)\| dx''.
\]

In Lemma 3.2.6 we showed that the right hand side of the above expression converges to 0. To obtain the claim, it is then enough to consider bounds on \(\int_{\mathbb{R}^d} |\psi(x'')| \|\partial_y p^0_1(y, x''; x)\| dx''\) and then translate them to \(\int_{\mathbb{R}^d} |\psi(x'')| \|\partial_y p^0_{1,\epsilon}(y, x''; x)\| dx''.\)

To that end, let us use first apply Bismut–Elworthy–Li Formula proved in Theorem 3.2.7 and second the non-negativity of \(\zeta^{x''}_\epsilon\), to obtain:

\[
\int_{\mathbb{R}^d} |\psi(x'')| \|\partial_y p^0_{1,\epsilon}(y, x''; x)\| dx'' = \int_{\mathbb{R}^d} |\psi(x'')| E \left[\zeta^{x''}_\epsilon (Y^{0,y,x}_1) \left(\int_0^1 (\Sigma^{-1}_t (r, Y^{0,y,x}_r, x) Z_s^0, y, x)_s^0 dB_s\right)^2\right] dx'' \\
\leq \int_{\mathbb{R}^d} E \left[|\psi(x'')|\zeta^{x''}_\epsilon (Y^{0,y,x}_1) \left(\int_0^1 (\Sigma^{-1}_t (r, Y^{0,y,x}_r, x) Z_s^0, y, x)_s^0 dB_s\right)^2\right] dx''
\]

Next, we take the limit as \(\epsilon \to 0\) just as in Theorem 3.2.7 and use the at most \(m\)th–order polynomial growth of \(\psi\), consequence of satisfying Definition 3.2.4 with \(m - 1\), and afterwards Young’s inequality together with Itô’s isometry, to obtain:

\[
\lim_{\epsilon \to 0} E \left[|\psi(x'')|\zeta^{x''}_\epsilon (Y^{0,y,x}_1) \left(\int_0^1 (\Sigma^{-1}_t (r, Y^{0,y,x}_r, x) Z_s^0, y, x)_s^0 dB_s\right)^2\right] \\
= E \left[|\psi(Y^{0,y,x}_1)| \left(\int_0^1 (\Sigma^{-1}_t (r, Y^{0,y,x}_r, x) Z_s^0, y, x)_s^0 dB_s\right)^2\right] \\
\leq C E \left[(1 + |Y^{0,y,x}_1|^m) \left(\int_0^1 (\Sigma^{-1}_t (r, Y^{0,y,x}_r, x) Z_s^0, y, x)_s^0 dB_s\right)^2\right] \\
\leq C \left[\left(\int_0^1 (\Sigma^{-1}_t (r, Y^{0,y,x}_r, x) Z_s^0, y, x)^2 ds\right) + (1 + E[|Y^{0,y,x}_1|^{2m}])\right].
\]

Finally, the fact that \(\Sigma^{-1}\) has growth of at most order \(m/2 - 1\) and we have uniform moment bounds for \(Y\) and \(Z\) (Lemmas B.1.3,B.1.6) allow us to conclude for some \(C > 0\):

\[
\int_{\mathbb{R}^d} |\psi(x'')| \|\partial_y p^0_1(y, x''; x)\| dx'' = \lim_{\epsilon \to 0} \left[|\psi(x'')|\zeta^{x''}_\epsilon (Y^{0,y,x}_1) \left(\int_0^1 (\Sigma^{-1}_t (r, Y^{0,y,x}_r, x) Z_s^0, y, x)_s^0 dB_s\right)^2\right] \\
\leq C(1 + |y|^{2m} + |x|^{2m}).
\]

Taking \(\psi\) to be a polynomial of order \(m\) (which implies that \(\phi \in \mathcal{S}_{2(m-1)}\)) and \(y = x\) gives us the second claim.

\[\square\]

### 3.2.2 Convergence to invariant measure

We devote this section to obtain Assumption 3.1.2 from Assumptions 3.2.1, 3.2.2 and 3.2.3. Because SDE (3.2.2) is actually defined through a McKean-Vlasov SDE such as (3.1.1), one can prove that they both have an invariant measure under Assumption 3.2.3. Moreover, since
\(X^\xi = Y_s^{0,\xi} \) a.s. by Lemma C.2.1, the process \((Y_s^{\xi,\lambda})_{s \geq 0}\) will converge to the same one and at the same rate.

The next thing to notice is that if \(\mathcal{F}\) is formed by those measurable functions which are bounded by some fix order polynomial family, Assumption 3.1.2 is satisfied with \(\lim_{s \to \infty} G(s) = 0\) once we have convergence to the invariant measure \((q)\) in Weighted Total Variation distance (see [3] and Chapter 2). However, here we consider the restriction of the family of evaluation functions to \(\mathcal{F}_m\) defined in (3.2.1), for which convergence to an invariant measure in \(m\)-Wasserstein distance is enough to obtain Assumption 3.1.2.

Next we provide a generalization of Proposition 4.3.4.

**Lemma 3.2.10.** Under Assumption 3.2.3, the process \((Y_s)_{s \geq 0}\) defined by (3.1.3) satisfies the following: there exist \(C, c_1, c_2 > 0\) such that for any \(2 \leq m \leq (p - 1)/2\), \(\xi, \xi', \chi, \chi' \in L^m(\mathbb{R}^d)\) and \(\tau \geq 1\),

\[
E[\|Y^\tau_{1+s} - Y^\tau_{1+s} X^\xi\|_m] \leq C(E|\chi - \chi'|^m e^{-c_1 s} + (1 + E|\xi - \xi'|^m e^{-c_2 s}), \quad \forall s \geq 0. \tag{3.2.19}
\]

Moreover, for any \(m \in \mathbb{N}\) such that \(2m < p\) and \(\xi \in L^m(\mathbb{R}^d)\) there exists a measure \(q = q^\xi \in \mathcal{P}_m(\mathbb{R}^d)\) such that:

\[
\mathcal{W}_m(\mathcal{L}(Y^\tau_{1+s} X^\xi), q(\cdot))_{|_{Y^\tau_{1+s} X^\xi}} = e^{-m M s} \mathcal{W}_m(\mathcal{L}(\xi), q), \quad \text{for all } s \geq 0.
\]

**Proof.** Let \(\lambda\) be an arbitrary positive constant, \(\xi, \xi', \chi, \chi' \in L^m(\mathbb{R}^d)\) starting random variables, where \(2m + 1 \leq p\). If Itô's formula is applied to the process \((e^{\lambda s} Y^\tau_{1+s} X^\xi - Y_{1+s}^\tau X^\xi - m s)^{m}\) \(s \geq t\), and after renaming

\[
\Delta \sigma \sigma^*(Y^\tau_{s} X^\xi, \mathcal{L}(X^\xi)), \mathcal{L}(X^\xi), \mathcal{L}(X^\xi)) = \left(\sigma(Y^\tau_{s} X^\xi, \mathcal{L}(X^\xi)) - \sigma(Y^\tau_{s} X^\xi, \mathcal{L}(X^\xi))\right) \left(\sigma(Y^\tau_{s} X^\xi, \mathcal{L}(X^\xi)) - \sigma(Y^\tau_{s} X^\xi, \mathcal{L}(X^\xi))\right)^*,
\]

we obtain the following:

\[
d(e^{\lambda s} Y^\tau_{1+s} X^\xi - Y^\tau_{1+s} X^\xi - m) = e^{\lambda s} \left(\lambda |Y^\tau_{1+s} X^\xi - Y^\tau_{1+s} X^\xi - m
+ m|Y^\tau_{s} X^\xi - Y^\tau_{s} X^\xi - m - 2(Y^\tau_{s} X^\xi - Y^\tau_{s} X^\xi, b(Y^\tau_{s} X^\xi, \mathcal{L}(X^\xi)) - b(Y^\tau_{s} X^\xi, \mathcal{L}(X^\xi))\right)
+ \frac{m(m-1)}{2} |Y^\tau_{s} X^\xi - Y^\tau_{s} X^\xi - m - 2 |Y^\tau_{s} X^\xi - Y^\tau_{s} X^\xi, b(Y^\tau_{s} X^\xi, \mathcal{L}(X^\xi)) - b(Y^\tau_{s} X^\xi, \mathcal{L}(X^\xi)))| ds
+ e^{\lambda s} \left(\lambda |Y^\tau_{s} X^\xi - Y^\tau_{s} X^\xi - m - 1 |\sigma(s, Y^\tau_{s} X^\xi, \mathcal{L}(X^\xi)) - \sigma(s, Y^\tau_{s} X^\xi, \mathcal{L}(X^\xi)))| dB_s.
\]

And now, by taking expectations of the above and using a stopping time argument as in Lemma B.1.3,

\[
dE[e^{\lambda s} Y^\tau_{1+s} X^\xi - Y^\tau_{1+s} X^\xi - m] = e^{\lambda s} \left(\lambda E[|Y^\tau_{1+s} X^\xi - Y^\tau_{1+s} X^\xi - m
+ mE[|Y^\tau_{s} X^\xi - Y^\tau_{s} X^\xi - m - 2(Y^\tau_{s} X^\xi - Y^\tau_{s} X^\xi, b(Y^\tau_{s} X^\xi, \mathcal{L}(X^\xi)) - b(Y^\tau_{s} X^\xi, \mathcal{L}(X^\xi)))]
+ \frac{m(m-1)}{2} |Y^\tau_{s} X^\xi - Y^\tau_{s} X^\xi - m - 2 |Y^\tau_{s} X^\xi - Y^\tau_{s} X^\xi, b(Y^\tau_{s} X^\xi, \mathcal{L}(X^\xi)) - b(Y^\tau_{s} X^\xi, \mathcal{L}(X^\xi)))| ds
+ e^{\lambda s} \left(\lambda E[|Y^\tau_{s} X^\xi - Y^\tau_{s} X^\xi - m - 1 |\sigma(s, Y^\tau_{s} X^\xi, \mathcal{L}(X^\xi)) - \sigma(s, Y^\tau_{s} X^\xi, \mathcal{L}(X^\xi)))| dB_s.
\]

\[71\]
which can be dealt with the monotonicity in Assumption 3.2.3 in order to obtain
\[ d\mathbb{E}[e^{\lambda s}|Y^T_{s} - Y^T_{s} - Y^{\tau,\xi}_{s} |^m] \leq e^{\lambda s}((\lambda - M_1 m)\mathbb{E}[|Y^T_{s} - Y^T_{s} |^m] + mM_0 \mathbb{E}[|X^\tau_{s} - X^{\tau,\xi}_{s}|^2]E|\mathbb{E}[|Y^T_{s} - Y^{\tau,\xi}_{s} |^m])ds. \]

(3.2.21)

In particular, if \( \xi = \xi' \), the inequality (3.2.22) becomes with the choice of \( \lambda = mM_1 > 0 \),
\[ d\mathbb{E}[e^{M_1 s}|Y^T_{s} - Y^T_{s} - Y^{\tau,\xi}_{s} |^m] \leq 0 ds, \]
which, on one hand, by integrating in between \( \tau \) and \( \tau + s \) implies
\[ \mathbb{E}[|Y^T_{\tau+s} - Y^T_{\tau+s} |^m] \leq e^{-M_1 s}(\mathbb{E}[|\xi - \xi'|^m]). \]

On the other hand, by the properties of the Wasserstein metric,
\[ W_m(\mathcal{L}(Y^T_{\tau+s}), \mathcal{L}(Y^T_{\tau+s})) \leq \mathbb{E}[|Y^T_{\tau+s} - Y^T_{\tau+s} |^m]. \]

Hence, (3.2.22) implies
\[ d\left(e^{M_1 s}W_m(\mathcal{L}(Y^T_{\tau+s}), \mathcal{L}(Y^T_{\tau+s}))\right) \leq 0 ds. \]

And if we now integrate in between \( \tau \) and \( \tau + s \),
\[ W_m(\mathcal{L}(Y^T_{\tau+s}), \mathcal{L}(Y^T_{\tau+s})) \leq e^{-M_1 s}\mathbb{E}[|\xi - \xi'|^m]. \]

Finally, since the Wasserstein metric is an infimum over all couplings and the solution of the McKean—Vlasov SDE does not depend on the random variable used as initial data, but rather it’s law, we conclude:
\[ W_m(\mathcal{L}(Y^T_{\tau+s}), \mathcal{L}(Y^T_{\tau+s})) \leq e^{-M_1 s}W_m(\mathcal{L}(\chi), \mathcal{L}(\chi')). \]

(3.2.23)

Next we show that this bound implies, for any \( \xi \in L^m(\mathbb{R}^d) \), convergence to invariant measure in \( m \)-Wasserstein distance for a process \( Y^T, X^\xi \). Indeed, let us define the map \( \mathcal{F}_s : P_m(\mathbb{R}^d) \rightarrow P_m(\mathbb{R}^d) \) by \( \mathcal{F}_s(\mu) = \mathcal{L}(Y^T_{\tau+s}, X^\xi) \), where \( \chi \sim \mu \). It is well defined (and, in particular, independent of the choice of \( \chi \)) and moreover a semigroup because of the existence and uniqueness solution to a McKean—Vlasov SDE (see [76, Theorem 6.1]). Additionally, fixed \( s_0 > 0 \) small enough such that \( e^{-M_1 s_0} < 1 \), from (3.2.23) we know that \( \mathcal{F}_s \) is a contraction in the complete metric space \( (P_m(\mathbb{R}^d), W_m) \). Consequently, by Banach’s Fixed Point Theorem, there exist a unique measure in \( P_m(\mathbb{R}^d) \), which we denote by \( \bar{\mu} = \bar{q}^\xi \), that is a fixed point for \( \mathcal{F}_{s_0} \). It is key to notice that \( \bar{q} \) might depend on \( s_0 \), which would be an impediment to extend the argument to all \( s \) and obtain existence and uniqueness of such an element in the whole
time interval $[0, \infty)$. Let us prove next that even if this is the case, we can construct another object which is in fact a fixed point for the whole family $(\mathcal{F}_s)_{s \geq 0}$.

Consider first $r \leq s_0$ and notice that the properties of the semigroup imply that the new measure $q = q^r := \int_0^{s_0} \mathcal{F}_s(q^r) \, ds$, still belonging to $\mathcal{P}_m(\mathbb{R}^d)$ (by Hölder’s inequality, (3.2.23)) and the fact that $\bar{q} \in \mathcal{P}_m(\mathbb{R}^d)$, satisfies:

$$
\mathcal{F}_r q = \mathcal{F}_r \int_0^{s_0} \mathcal{F}_s(q) \, ds = \int_0^{s_0} \mathcal{F}_r \mathcal{F}_s(q) \, ds = \int_0^{s_0} \mathcal{F}_{s+r}(q) \, ds = \int_r^{s_0+r} \mathcal{F}_s(q) \, ds
$$

Recalling now that $\bar{q}$ is a fixed point of $\mathcal{F}_{s_0}$, we conclude that indeed $q$ is a fixed point for $\mathcal{F}_r$ for all $0 \leq r \leq s_0$:

$$
\mathcal{F}_r q = \mathcal{F}_{s_0} \mathcal{F}_{r-s_0} q = (\mathcal{F}_{s_0})^k \mathcal{F}_{r-s_0} q = q.
$$

Next consider the possibility of $s_0 < r$, in which case there exists a natural number $k$ such that $k \cdot s_0 < r \leq (k+1) \cdot s_0$. Hence, notice that a simple iteration allows us to step on the previous case since $r - k \cdot s_0 \leq s_0$:

$$
\mathcal{F}_r q = \mathcal{F}_{k \cdot s_0} \mathcal{F}_{r-k \cdot s_0} q = (\mathcal{F}_{s_0})^k \mathcal{F}_{r-k \cdot s_0} q = q.
$$

We prove therefore that there must exist a fixed point for $(\mathcal{F}_s)_{s \geq 0}$, which in other words means an invariant measure for the process $(X_{T+s}^T, X^\xi_{T+s}^T)$, this together with (3.2.23) give the bound in our statement. Indeed, if we choose $\chi'$ distributed as $q$ in (3.2.23)

$$
\mathcal{W}_m(\mathcal{L}(X_{T+s}^T, q), \mathcal{L}(Y_{T+s}^T, X^\xi_{T+s}^T)) \leq e^{-M_1 s} \mathcal{W}_m(\mathcal{L}(\chi), q).
$$

Moreover, when $\chi$ is taken to be distributed as an invariant measure with $\mathcal{L}(\chi) \neq q$, from (3.2.24)

$$
\mathcal{W}_m(\mathcal{L}(\chi), q(\cdot)) = \mathcal{W}_m(\mathcal{L}(X_{T+s}^T, q), \mathcal{L}(Y_{T+s}^T, X^\xi_{T+s}^T)) \leq e^{-M_1 s} \mathcal{W}_m(\mathcal{L}(\chi), q).
$$

But since this must hold for all $s \geq 0$, we conclude that $\chi' \sim \chi$, i.e. uniqueness of the invariant measure $q$.

Another particular case to consider is $\chi = \chi' = \xi$. In this scenario, given that $X_{T+s}^T = Y_{T+s}^T, X^\xi_{T+s}^T$ (see Lemma C.2.5), equality (3.2.21) and monotonicity in Assumption 3.2.3 imply in the same way that

$$
\mathbb{E}[|X_s^\xi - X_s^\xi|^{m}] \leq CE|\xi - \xi'|^{m} e^{-m(M_1-M_0)s}.
$$

Going back now to (3.2.22), by Young’s inequality with an arbitrary $c > 0$ and (3.2.25), the following is satisfied for some $C_\epsilon, \epsilon > 0$:

$$
d\mathbb{E}[e^{\lambda s} |Y_{T+s}^T - Y_{T+s}^T, X^\xi_{T+s}^T|^{m}] \leq e^{\lambda s}(\lambda - M_1 m + me)\mathbb{E}[Y_{T+s}^T, X^\xi_{T+s}^T - Y_{T+s}^T, X^\xi_{T+s}^T|^{m} + C_\epsilon(1 + e^{-c s}\mathbb{E}[|\xi - \xi'|^{m}])] \, ds
$$

Now, for the particular choice of $\lambda = m(M_1 - \epsilon)$, where we can pick $\epsilon$ small enough such that $\lambda > 0$,

$$
d\mathbb{E}[e^{m(M_1-\epsilon)s} |Y_{T+s}^T - Y_{T+s}^T, X^\xi_{T+s}^T|^{m}] \leq C_\epsilon(e^{m(M_1-\epsilon)s} + e^{-(c-m(M_1-\epsilon))s}\mathbb{E}[|\xi - \xi'|^{m}]) \, ds,
$$

73
which by integrating in between \( \tau \) and \( s \) becomes for a different values of \( c, C_c > 0 \),
\[
\mathbb{E} \left[ Y_{T_{\tau+s}}^{T,\xi} - Y_{T_{\tau}}^{T,\xi} \right] = m \leq e^{-m(M_1 - c)s} \mathbb{E}|\chi - \chi'|^{m} + C_c(1 + e^{-c_3\mathbb{E}})|\xi - \xi'|^{m})
\]
After renaming our constants, we derive (3.2.19).

\[\square\]

**Theorem 3.2.11.** Given \( m \geq 2 \), let Assumptions 3.2.1, 3.2.2, 3.2.3 hold and consider functions \( \phi \in \mathcal{F}_m \) defined in (3.2.1) for \( m \leq (p - 1)/2 \). For any \( s \geq 0, x, y \in \mathbb{R}^d \), recall that \( p_{1+s}^{1+s}(y, \cdot; x) \) denotes the density of \( Y_{1+1}^{1,\xi,x} \) solving 3.1.1 and \( q^x \) its invariant measure. Then there exist \( C, c > 0 \) such that:
\[
\sup_{\phi \in \mathcal{F}_m} \left| \int_{\mathbb{R}^d} \phi(x') p_{1+s}^{1+s}(y, dx'; x) - q^x(dx') \right| \leq C(1 + |x|^m + |y|^m) e^{-cs}, \quad \forall (s, x) \in [0, \infty) \times \mathbb{R}^d.
\]
In other words, Assumption 3.1.2 is satisfied with \( G(s) = e^{-cs} \) and \( g(x) = C(1 + |x|^m) \) for some constants \( C, c > 0 \), which might depend on the invariant measure. Nonetheless, Assumption 3.1.2 still holds since the invariant measure is fixed for Theorem 3.1.4.

**Proof Theorem 3.2.11.** Fix \( s \geq 0 \) and recall that \( q^x \) denotes the invariant measure of the process \( (Y_{1+s}^{1,\xi,x})_{t \geq 0} \) whose existence we proved in Proposition 3.2.10. Then, for any coupling \( \pi \) between \( p_s \) and \( q \), and for any \( \phi \in \mathcal{F}_m \),
\[
\left| \int_{\mathbb{R}^d} \phi(x') p_{1+s}^{1+s}(y, dx') - \int_{\mathbb{R}^d} \phi(x') q^x(dx') \right| \leq \int_{\mathbb{R}^d} \left| \phi(x') - \phi(x'') \right| \int_{\mathbb{R}^d} \pi(dx', dx'')
\]
\[
\leq C \int_{\mathbb{R}^d} \left| x' - x'' \right| \int_{\mathbb{R}^d} \pi(dx', dx'') + C \left( \int_{\mathbb{R}^d} \left| x'^m p_{1+s}^{1+s}(y, x'; x) + \left( \int_{\mathbb{R}^d} \pi(dx', dx'') \right)^{1/2} \left( \int_{\mathbb{R}^d} \left| x' - x'' \right|^2 \pi(dx', dx'') \right)^{1/2} \right) \right)
\]
Finally, taking infimum over such couplings and applying Lemma 3.2.10 together with the fact that \( q^x \in \mathcal{P}_m(\mathbb{R}^d) \) and must satisfy the same moments estimates as \( p_s(\cdot, \cdot; x) \) (due to the contraction), we conclude that there exist \( C, c > 0 \) such that:
\[
\left| \int_{\mathbb{R}^d} \phi(x') p_{1+s}^{1+s}(y, dx') - \int_{\mathbb{R}^d} \phi(x') q^x(dx') \right| \leq C \mathcal{W}_1 (\mathcal{L}(Y_{1+s}^{1,\xi,x}, q^x) + C \mathcal{W}_2^{1/2} (\mathcal{L}(Y_{1+s}^{1,\xi,x}, q^x)) \mathbb{E}|Y_{1+s}^{1,\xi,x}|^{m}^{1/2} \left( \int_{\mathbb{R}^d} \left| x'^m q^x(dx') \right|^{1/2} \right)
\]
\[
\leq C(1 + |x|^m + |y|^m) e^{-cs}.
\]
In other words, since the above was holding for arbitrary test functions in the family \( \mathcal{F}_m \),
\[
\sup_{\phi \in \mathcal{F}_m} \left| \int_{\mathbb{R}^d} \phi(x') (p_s^0(y, dx'; x) - q^x(dx')) \right| \leq C(1 + |x|^m + |y|^m) e^{-cs}.
\]
\[\square\]
Remark 3.2.12. We stress the fact that throughout the process of using Theorem 3.1.4 to bound $\partial_y V(t, x; x)|_{y=x}$, we never rely on the regularity of the test function. However, if the family of test functions $\mathcal{F}$ can in fact be embedded within functions belonging to $C^2(\mathbb{R}^d)$, such that itself and its derivatives are polynomially growing, then we could directly obtain derivatives decay for $V$ using a combination of the Chain Rule and the decaying moments bounds for the derivative process (see Chapter 4). Moreover, this is easily generalized to higher order derivatives.

3.3 Derivative with respect to the parameter of the auxiliary backward Kolmogorov function

Recall that we define $(s, x) \mapsto V(s, x) = \mathbb{E}[\phi(X^x_s)]$, for $(X^x_s)_{s \geq 0}$ solution to the SDE (3.1.1) and $\phi \in \mathcal{F}_m$ (see (3.2.1)). Additionally, recall that we defined the function $V(s, y; x) = \mathbb{E}[\phi(Y^{0,y,x}_s)]$, where $(Y^{0,y,x}_s)_{s \geq 0}$ is the solution to the SDE (3.2.2). Hence, $\partial_s V(s, x) = \partial_y V(s, y; x)|_{y=x} + \partial_x V(s, y; x)|_{y=x}$. We already know from Theorem 3.2.5 that, under Assumptions 3.2.1, 3.2.2 and 3.2.3, $V$ satisfies for some $C, c > 0$ the following:

$$|\partial_y V(s, y; x)|_{y=x} \leq C(1 + |x|^m) e^{-cs}, \quad \text{for all } (s, x) \in [1, \infty) \times \mathbb{R}^d. \quad (3.3.1)$$

For completeness, in Theorem 3.3.4 we obtain, if not decaying, at least explicit bounds on $|\partial_x V(s, y; x)|_{y=x}$.

Recall that $(s, x') \mapsto p^0_s(y, x'; x)$ denotes the density of $Y^{0,y,x}_s$. Notice that if we had an integrable and positive function bounding $|\phi(\cdot)\partial_x p^0_s(y, \cdot; x)|$ uniformly in $x$ for any fixed $s > 0$, $y \in \mathbb{R}^d$, we could apply the well known corollary of the Lebesgue Dominated Convergence Theorem for differentiating under the integral sign to obtain:

$$\partial_x V(s, y; x) = \partial_x \int_{\mathbb{R}^d} \phi(x') p^0_s(y, x'; x) dx' = \int_{\mathbb{R}^d} \phi(x') \partial_x p^0_s(y, x'; x) dx', \quad \forall x, y \in \mathbb{R}^d, s \geq 0.$$  

Our strategy for bounding $\int_{\mathbb{R}^d} \phi(x') \partial_x p^0_s(y, x'; x) dx'$ is based on a characterization of the derivative of transition density with respect to the parameter $x$ in terms of the density itself and a second order operator's action on it (the derivatives are in $x'$ this time; see Proposition 3.3.3). Moreover, although necessary for technical details, decaying bounds of $(s, x') \mapsto p^0_s(y, x'; x)$ and its derivatives ($(s, x') \mapsto \partial_{x'} p^0_s(y, x'; x)$ for $i = 1, 2$), which are integrable over $\mathbb{R}^d$ against $\phi$ are not directly applicable. This direct computation does not result in finite estimates. Instead, we perform integration by parts and use decaying moments estimates for $Y^{0,y,x}_s$ to obtain explicit bounds for $|\partial_x V(s, y; x)|_{y=x}$ in Theorem 3.3.4.

Next we pose assumptions on the coefficients of the second order operator, which we announced: appears in the characterization of $\partial_x p^0_s(y, x'; x)$.

Assumption 3.3.1. For any $t \geq 0$, the coefficients $\mathbb{R}^d \times \mathbb{R}^d \ni (y, x) \mapsto b(y, \mathcal{L}(X^x_t)), \sigma(y, \mathcal{L}(X^x_t))$ are Lipschitz in the measure component and differentiable enough such that the following is satisfied: there exist $C > 0, p_1 \geq 0$ such that for all $x, y \in \mathbb{R}^d$,

$$\sup_{t \geq 0} \max \left\{ |\partial_x \partial_{x'}^i b(y, \mathcal{L}(X^x_t))|, |\partial_x \partial_{x'}^j \sigma \sigma^* (y, \mathcal{L}(X^x_t))| : i = 0, 1; j = 0, 1, 2 \right\} \leq C(1 + |x|^{p_1} + |y|^{p_1}).$$
Notice that our argument, consisting in obtaining explicit estimates for the derivatives of the backward Kolmogorov function $V$, based on already requiring uniform differentiability of the coefficients in the initial datum of the law component, is not necessarily cyclic. Naturally, if one had uniform derivatives (in $x$) of the transition density $p(\cdot,\cdot; x)$, i.e., solution to the backward Kolmogorov equation, this translates immediately to uniform derivative estimates for $\partial_x V$, given that the test function $\phi$ is continuous. In that case, Assumption 3.3.1 holds but at the same time there is no need for the extra work presented in this chapter in order to obtain polynomial growth of $\partial_x V$.

So, without loss of generality, let us pick the coefficient $b(y, \mathcal{L}(X_t^x))$ and explain what we mean by differentiating it with respect to $x$ and the fact that is polynomially growing. The rest of the terms in Assumption 3.3.1: derivative in $x$ of $\partial_y b(y, \mathcal{L}(X_t^x)), \partial_y \sigma \sigma^* (y, \mathcal{L}(X_t^x))$ with $j = 0, 1, 2$, follow by applying exactly the same analysis.

One way of bounding $x \mapsto b(y, \mathcal{L}(X_t^x))$ uniformly in $y, t$ is the following. Observe that under the appropriate regularity assumptions and due to the Chain Rule for measure–dependent functions (see [35, Theorem 5.92]), the following is satisfied:

$$
\partial_x b(y, \mathcal{L}(X_t^x)) = \mathbb{E}[\partial_\mu b(y, \mathcal{L}(\tilde{X}_t^x))(\tilde{X}_t^x) \partial_x (\tilde{X}_t^x)],
$$

for $\tilde{X}_t^x$ a copy of $X_t^x$ in a independent probability space where we denote expectations by $\tilde{E}$; and where $\partial_x \tilde{X}_t^x$ is a well defined process (see [35]). Hence, due to the Cauchy–Schwarz inequality, if we bound uniformly (in $y$ and $t$) the second moments of $(\partial_x \tilde{X}_t^x)_{t \geq 0}$ as well as $\partial_\mu b(y, \mathcal{L}(\tilde{X}_t^x))(\tilde{X}_t^x)$, we obtain uniform (in $y$ and $t$) bounds for $\partial_x b(y, \mathcal{L}(X_t^x))$. On one hand, the second moment of $\partial_x \tilde{X}_t^x$ can be bound with a similar trick to that in Chapter 4, which is based on the monotonicity in Assumption 3.2.1, only for the derivative of a McKean–Vlasov processes instead. On the other hand, the bound on the second moment of $\partial_\mu b(y, \mu)|_{\mu = \mathcal{L}(X_t^x)}$ can be obtained, for example, from requiring that the coefficients of the original McKean–Vlasov SDE have bounded measure derivative. An example of a drift satisfying such a condition is $b(x, \mathcal{L}(X_t^x)) = \mathbb{E}[\beta(X_t^x)]$ for some function $\beta : \mathbb{R}^d \to \mathbb{R}^d$ with bounded derivative $\partial_x \beta$. Indeed, it is known that in this case $\partial_\mu b(x, \mathcal{L}(X_t^x))(\tilde{X}_t^x) = (\partial_x \beta)(\tilde{X}_t^x)$, which is assumed bounded. This is an approach closer to Chapter 4, which can be studied in the future.

However, this is not the only way to obtain Assumption 3.3.1. Indeed, suppose that we only know that $\partial_x p(\cdot,\cdot; x)$ exists (without any extra requirements on its growth) and that $b$ is differentiable in the measure component, which imply that $x \mapsto b(y, \mathcal{L}(X_t^x))$ exists. Suppose moreover that Assumption 3.2.1 holds and that, fixed $2 \leq p_1 < p$ and $b$, there exists $C_b > 0$ such that for any $t \geq 0, x, x_0, y, y_0, \in \mathbb{R}^d$,

$$
|b(y, \mathcal{L}(X_t^x)) - b(y_0, \mathcal{L}(X_t^{x_0}))| \leq C_b (1 + |y|^{p_1/2} + |y_0|^{p_1/2})|y - y_0| + C_b (1 + \mathbb{E}[|X_t^x|^{p_1/2}] + \mathbb{E}[|X_t^{x_0}|^{p_1/2}])\mathcal{W}_1(\mathcal{L}(X_t^x), \mathcal{L}(X_t^{x_0})).
$$

(3.3.2)

We could name this condition as $p_1$–locally Lipschitz in both components (in the measure component with respect to the 1–Wasserstein distance).

Since by the bounds obtained in (3.2.25) and Lemma B.1.2 there exist $C, c > 0$ such that
for all \( x, x_0, t \geq 0 \),
\[
\mathcal{W}_t(\mathcal{L}(X_t^x), \mathcal{L}(X_t^{x_0})) \leq \left( \mathbb{E}[|X_t^x - X_t^{x_0}|^2] \right)^{1/2} \leq C e^{-ct} |x - x_0| \quad \text{and} \quad \mathbb{E}[|X_t^x|^{p_1/2}] \leq C e^{-ct} (1 + |x|^{p_1/2}),
\]
we continue bounding in (3.3.2):
\[
|b(y, \mathcal{L}(X_t^x)) - b(y, \mathcal{L}(X_t^{x_0}))| \leq C_b(1 + |y|^{p_1/2} + |y_0|^{p_1/2}) |y - y_0| + C_b C e^{-ct} (1 + |x|^{p_1} + |x_0|^{p_1}) |x - x_0|.
\]
Let us fix next \( y \in \mathbb{R}^d \) and an orthonormal basis of \( \mathbb{R}^d \): \( \{e_1, ..., e_d\} \), \( h > 0 \) and an index \( i \in \{1, ..., d\} \). Then for \( x_0 = x + he_i \) in the above expression,
\[
\lim_{h \to 0} \frac{b(y, \mathcal{L}(X_t^x)) - b(y, \mathcal{L}(X_t^{x+he_i}))}{h} \leq \frac{C_b \lim_{h \to 0} \left( 1 + |y|^{p_1/2} + |y|^{p_1/2} \right) |y - y| + C(1 + |x|^{p_1} + |h|^{p_1}) h}{h} \leq CC_b(1 + |x|^{p_1}),
\]
or in other words: if the derivative \( \partial_x b((y, \mathcal{L}(X_t^x))) \) exists, it must have polynomial growth in \( x \) of order \( p_1 \), uniformly in \( y \) and Assumption 3.3.1 is achieved.

**Assumption 3.3.2.** The coefficients \( b, \sigma \in \mathcal{C}^{4,0} (\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)) \) are uniformly bounded in the measure component. Moreover, for any \( t \geq 0 \), the function \( \mathbb{R}^d \times \mathbb{R}^d \ni (y, x) \mapsto 1/2 \sigma \sigma^*(y, \mathcal{L}(X_t^x)) \) is bounded. Finally, the following is satisfied: there exist \( C, M_2, \epsilon > 0 \) such that for \( |k| = 1, 2 \),
\[
\sup_{\mu} |\partial_{\mu}^k \sigma \sigma^*(y, \mu)| \leq C(1 + |y|)^{|k|(1-\epsilon)};
\]
\[
\sup_{\mu} |b(y, \mu)| \leq C(1 + |y|) \quad \text{and} \quad \sup_{\mu} |\partial_{\mu}^k b(y, \mu)| \leq C(1 + |y|)^{1+|k|(1-\epsilon)};
\]
\[
\sup_{\mu} \{|\partial_{y}^j k b(y, \mu)| + |\partial_{y}^2 k \sigma \sigma^*(y, \mu)|\} \leq C(1 + |y|)^{2+|k|(1-\epsilon)};
\]
\[
\sup_{\mu} \{|\partial_{y}^j b(y, \mu)| + \partial_{y}^2 \sigma \sigma^*(y, \mu)| \leq -M_2(1 + |y|)^2.
\]

Under Assumption 3.3.2 and uniform ellipticity, we recall once more the conclusion to Eidelman’s theorems (see [80, Theorem 2])
\[
|\partial_{x^n}^j p_t^T(y, x^n; x)| \leq C s^{-(d+j)/2} e^{\frac{1}{2} \delta (y - x''^n)^2}, \quad \forall y, x'' \in \mathbb{R}^d, j = 0, 1, 2. \tag{3.3.3}
\]

The quadratic exponential decay of the transition density and its first two space derivatives can be obtained under different set of assumptions (see e.g [25, Theorem 9.4.2, Remark below display (9.4.18)], [26, Theorem VI.2, VI.5] or [68]).

Let us mention that the condition (3.3.3) is stronger than what we need (for the the transition density to integrate a \( p_1 \)-order polynomial and existence of its two space derivatives) and we only introduce it here in order to see the link in between Chapter 2 and the present work.

An intermediate result and key characterization of \( \partial_x p_t^T(y, x'; x) \), of interest on its own and inspired by [28], is the following.
Proposition 3.3.3. Fixed \( x, y \in \mathbb{R}^d, \tau \geq 0 \), consider \((t, x') \rightarrow p^T_t(y, x'; x)\) density of the process \((Y^x_{\tau}))_{s \geq 0}\) defined by (3.2.2). Then, under Assumptions 3.2.2, 3.3.1 and 3.3.2 the following is satisfied: for all \( x' \in \mathbb{R}^d, (\tau, \infty) \times \mathbb{R}^d \ni (t, x) \rightarrow \partial_x p^T_t(y, x'; x)\) exists, is continuous and satisfies the following:

\[
\partial_x p^T_t(y, x'; x) = \int_{\mathbb{R}^d} \int_t^T \partial_{x-s}(y, x'-x''; s)(\partial_x L^x)(s, x''; x) p^T_s(y, x''; x) dx'' ds. \tag{3.3.4}
\]

Proof. Let us assume for this proof, without loss of generality, that \( \tau = 0 \). Notice that fixed \( s \in (0, \infty) \) and given estimates (3.3.3) on \( p^0_t(y, x''; x), \partial_x^0 p^0_t(y, x''; x) \) and \( \partial_{x''}^0 p^0_t(y, x''; x) \), together with the Assumption 3.3.1 (which also makes \( \partial_x L^x \) well defined) we know that the integrability over \( dx'' \) in (3.3.4) is no issue.

Let us see first how we derive formula (3.3.4). Recall that \((t, x'') \to p^0_t(y, x''; x)\) satisfies the following (forward Kolmogorov) equation: for \( x, y \in \mathbb{R}^d \) fixed and also a fixed terminal time \( T > 0 \),

\[
\begin{cases}
\partial_t p^0_t(y, x''; x) - L^x(t, x''; x) p^0_t(y, x''; x) = 0, & (t, x'') \in (0, T] \times \mathbb{R}^d, \\
p^0_0(y, x''; x) = \delta_y(x''), & x'' \in \mathbb{R}^d.
\end{cases} \tag{3.3.5}
\]

Next we define a new object for \( h \neq 0 \) and \( e \) an arbitrary element of a fixed orthonormal basis \( \{e_i\}_{i=1,\ldots,d} \) in \( \mathbb{R}^d \) as follows:

\[
q^h_t(y, x''; x) := \frac{p^0_t(y, x''; x + he) - p^0_t(y, x''; x)}{h},
\]

which satisfies, heuristically derived from (3.3.5), the following: \( q^0_0(y, x''; x) = 0 \) and for all \( (t, x'') \in (0, T] \times \mathbb{R}^d \),

\[
\begin{aligned}
\partial_t q^h_t(y, x''; x) &= \frac{L^x(t, x''; x + he) p^0_t(y, x''; x + he) - L^x(t, x''; x) p^0_t(y, x''; x)}{h} \\
&= \frac{L^x(t, x''; x) p^0_t(y, x''; x + he)}{h} - \frac{p^0_t(y, x''; x)}{h} + f^{y,x}_h(t, x''),
\end{aligned} \tag{3.3.6}
\]

where \( f^{y,x}_h(t, x'') := \frac{L^x(t, x''; x + he) - L^x(t, x''; x)}{h} p^0_t(y, x''; x + he) \). Let us justify the meaning of the equation in (3.3.6). Because of the singularity at 0, it only makes sense to consider a solution to (3.3.6) when we discuss it in the weak sense, i.e. integrating against a smooth function \( \Gamma \) with compact support: for all \( t \in (0, T] \),

\[
\begin{aligned}
\int_{\mathbb{R}^d} \Gamma(z) \partial_t q^h_t(y, z; x) dz &= \int_{\mathbb{R}^d} \Gamma(z) \left[ L^x(t, y; x) q^h_t(y, z; x) + f^{y,x}_h(t, z) \right] dz, \\
\int_{\mathbb{R}^d} \Gamma(z) q^h_t(y, z; x) dz &= 0. \tag{3.3.7}
\end{aligned}
\]

So let us construct a sequence of these test functions \( \Gamma_k \) converging to \( \delta_{x''} \) as \( k \to \infty \). We see then that the estimates of the transition density’s derivatives \( \partial_x, \partial_x^2 \) and Assumption 3.3.1 for the coefficients, allow us to use Lebesgue Dominated Convergence Theorem in order to conclude that (3.3.6) has a solution in the sense of (3.3.7).

Moreover, since \( p^0_t(y, \cdot; x) \) is a fundamental solution of the first part of (3.3.5), the solution to the same PDE but with a certain forcing term (the forcing term being \( f^{y,x}_h(t, x'') \) in (3.3.6))
can be expressed using Duhamel’s principle (see [2, Chapter 2]) as

\[ q^h_t(y, x'; x) = \int_0^t \int_{R^d} p_{t-s}^0(y, x' - x''; x) f_{h}^{y,x}(s, x'') d x'' ds. \] (3.3.8)

Now, all that is left is to justify the pass to the limit as \( h \to 0 \) in (3.3.8) in order to obtain the claimed expression (3.3.4). This can be done with Leibniz integral rule for differentiating under the sign of the integral (see [18, Theorem 12.14]) as long as \( p_{t}^{0}(y, x', x) \), \( f_{h}^{y,x}(s, x''; x) \) (for any \( h > 0 \)) and \( \lim_{h \to 0} f_{h}^{y,x}(s, x''; x) = \partial_{x} L^{*}(s, x''; x) p_{t}^{0}(y, x''; x) \) are continuous in \( (x'', x) \) for any \( s \in [0, t] \). Recalling Assumptions 3.3.1–3.3.2 and in particular the continuity in \( x'' \) of the density (see [5]) we have all continuities in \( x'' \). It is then enough to check the continuity in the \( x \) component of \( p_{t}^{0}(y, z; x), \int_{R^d} \Gamma(z) \partial_{z} p_{t}^{0}(y, z; x) dz \) and \( \int_{R^d} \Gamma(z) \partial_{z}^{2} p_{t}^{0}(y, z; x) dz \) for any \( s, y, z \). The first continuity is consequence of (3.3.5) by the following argument: for any \( x, x' \in \mathbb{R}^{d} \),

\[
\partial_{t} p_{t}^{0}(y, x''; x) - \partial_{t} p_{t}^{0}(y, x''; x') = L^{*}(t, x''; x)(p_{t}^{0}(y, x''; x) - p_{t}^{0}(y, x''; x'))
\]

\[
+ (L^{*}(t, x''; x) - L^{*}(t, x''; x')) p_{t}^{0}(y, x''; x')
\]

and all the coefficients in \( L^{*} \) are continuous in \( x \) by Assumption 3.3.1. This, together with the fact that the identically zero function is the unique classical solution to the PDE stated for any \( (t, x'') \in (0, T) \times \mathbb{R}^{d} \) as

\[
\partial_{t} u(t, x'') = L^{*}(t, x'') u(t, x''); \quad u(0, x'') = 0,
\]

allows us to conclude that as \( |x - x'| \to 0, |p_{t}^{0}(y, x''; x) - p_{t}^{0}(y, x''; x')| \to 0. \)

On the other hand, continuity in \( x \) of \( \int_{R^d} \Gamma(z) \partial_{z} p_{t}^{0}(y, z; x) dz \) is consequence of the previous one by the following argument. Provided that \( \Gamma \in C^{0}((\mathbb{R}^{d}) \) (so we can apply integration by parts) and because of the continuity in \( z \) of \( p_{t}^{0}(y, z; x) \), we then have \( \partial_{z} p_{t}^{0}(y, z; x_{n}) \to \partial_{z} p_{t}^{0}(y, z; x) \) whenever \( x_{n} \to x \) up to a countable dense set of \( z \)'s. Another iteration gives us the continuity of \( \int_{R^d} \Gamma(z) \partial_{z}^{2} p_{t}^{0}(y, z; x) dz \) as well for \( \Gamma \in C^{2}(\mathbb{R}^{d}) \).

Finally, continuity of \( \partial_{x}^{2} p_{t}^{0}(y, x'; x) \) in \( x \) follows from uniform convergence of (3.3.8) to (3.3.4).

In the next theorem we obtain, if not decaying, at least explicit bounds for \( |\partial_{x} V(t, y; x)|_{y=x} \).

**Theorem 3.3.4.** Fix a natural number \( m \geq 2 \) and let Assumptions 3.2.1,3.2.2,3.3.1, 3.3.2 and property (B.1.1) hold with \( p \geq m + 2p_{1} + 2 \). Consider \( (Y_{s})_{s \geq 0} \) the stochastic process solution to (3.2.2). Given a function \( \phi \in C_{m}^{0} \) defined by (3.2.1), the function

\[
(0, \infty) \times \mathbb{R}^{d} \ni (s, x) \to V(s, y; x) = E[\phi(Y_{s}) \mid y] \quad \text{satisfies for some } C, \ c > 0 \text{ the following:}
\]

\[
|\partial_{x} V(s, y; x)|_{y=x} \leq C\left(M_{2}^{2} s + M_{2} |x|^{\max\{2, p_{1}\}} + |x|^{m/2 + 1} + M_{2} e^{-r t} + p_{1} e^{-c s}\right), \quad \text{for all } (s, x) \in (0, \infty) \times \mathbb{R}^{d}.
\]

Notice that if the coercivity property (B.1.1) holds with \( M_{2} = 0 \) and is not obtained by Lemma B.1.1 as a mere consequence of monotonicity condition in Assumption 3.2.3, these estimates are decaying in time.

**Proof Theorem (3.3.4).** We use first Proposition 3.3.3, second Fubini’s Theorem to exchange integration in the space variables, next Assumption 3.3.1 on \( \alpha_{0} := \partial_{x}(\partial_{y} \beta + \partial_{y}^{2} \text{tr}(\Sigma \Sigma^{*})) \), \( \alpha_{1} :=
In this section we gather our results and present estimates for the full gradient of $V^3$.4 Gradient of the backward Kolmogorov function for a

\[ \partial_x (\beta + \partial_x \text{tr}(\Sigma^*)) \] and $\alpha_2 := \partial_x (\Sigma^*)$ to obtain for all $t \geq 0, x, y \in \mathbb{R}^d$:

\[
\left| \partial_x V(t, y; x) \right|_{y=x} = \left| \int_{\mathbb{R}^d} \phi(x') \partial_x p^0_t(y, x'; x) |_{y=x} d x' \right|
\leq \int_{\mathbb{R}^d} \phi(x') \int_0^t \int_{\mathbb{R}^d} \partial_x p^0_{t-s}(y, x'; x) |_{y=x} d x'' d s d x' \]

and afterwards we use a change of variables, $(m/2 + 1)$–order polynomial order growth of test functions and the decay of the moments proved in Lemma B.1.2 to obtain:

\[
\leq C \int_{\mathbb{R}^d} \int_0^t (1 + |x''|^p_1 + |x|^p_1) \left( M_2 + e^{-c(t-s)} (|x|^{m/2+1} + |x''|^{m/2+1}) \right) \times (p^0_s(y, x''; x) |_{y=x} + |\partial x'' p^0_s(y, x''; x) |_{y=x} + |\partial^2 x'' p^0_s(y, x''; x) |_{y=x}) ds d x''.
\]  

Next notice that by integration by parts against the $p^1$–order polynomial (possible because of Aronson estimates at each fixed time $s$) and decay of the moments proved in Lemma B.1.2

\[
\int_{\mathbb{R}^d} (1 + |x''|^p_1 + |x|^p_1) |\partial x'' p^0_s(y, x''; x) |_{y=x} d x'' \leq C(M_2 + (|x|^2 + |x|^p_1-1) e^{-cs});
\]

and similarly, by integration by parts twice against the $p^1$–order polynomial and again decay of the moments,

\[
\int_{\mathbb{R}^d} (1 + |x''|^p_1 + |x|^p_1) |\partial^2 x'' p^0_s(y, x''; x) |_{y=x} d x'' \leq C(M_2 + (|x|^2 + |x|^p_1-2) e^{-cs}).
\]

Notice that one can always take $p_1$ greater in Assumption 3.3.1 and in particular work with $p_1 \geq 2$ for this step.

These estimates and the similar ones for the polynomials of degree $m/2 + p_1 + 1$, when taken back into (3.3.9), together with again decay of the moments used on the first summand, imply

\[
\left| \partial_x V(t, y; x) \right|_{y=x} \leq C \int_0^t \left( M_2 (M_2 + |x|^{\max[2, p_1]} e^{-cs}) + (M_2 + |x|^{m/2+1+p_1} e^{-cs}) e^{-c(t-s)} \right) ds
\leq C \left( M_2 t + M_2 |x|^{\max[2, p_1]} + M_2 e^{-ct} + |x|^{m/2+1+p_1} e^{-ct} \right).
\]

\[ \square \]

### 3.4 Gradient of the backward Kolmogorov function for a McKean–Vlasov SDE

In this section we gather our results and present estimates for the full gradient of $V(s, x) = \mathbb{E}[\phi(X^t_s, x)]$, where $X$ solves a McKean–Vlasov SDE.
**Corollary 3.4.1.** Let Assumptions 3.2.1, 3.2.2, 3.2.3, 3.3.1, 3.3.2 hold with \( p > \max(2m + 5, m + 2p_1 + 2) \). For any \( x \in \mathbb{R}^d \), consider \((X_s^x)_{s \geq 0}\), the stochastic process solution to (3.1.1). Given \( \phi \in S_m \) (defined in (3.2.1)) the function \((0, \infty) \times \mathbb{R}^d \ni (s, y) \mapsto V(s, x) = \mathbb{E}[\phi(X_s^0, x)]\) satisfies the following for some \( C, c > 0 \):

\[
|\partial_x V(s, x)| \leq C \left(M_2^2 s + M_2 |x|^{\max(2, p_1)} + (M_2 + |x|^{\max(2m, m/2 + 1 + p_1)}) e^{-cs}\right), \quad \text{for all } (s, x) \in (1, \infty) \times \mathbb{R}^d.
\]

In particular, if property (B.1.1) holds with \( p \geq m + 2p_1 + 2 \) and \( M_2 = 0 \),

\[
|\partial_x V(s, x)| \leq C(1 + |x|^{\max(2m, m/2 + 1 + p_1)}) e^{-cs}, \quad \text{for all } (s, x) \in [1, \infty) \times \mathbb{R}^d.
\]

**Proof.** It is a direct application of Theorems 3.2.5 and 3.3.4 and decomposition (3.2.3). \( \square \)

**Example 3.4.2.** We present here an example of a process solution to a McKean Vlasov SDE and such that the solution to the corresponding backward Kolmogorov equation has a space derivative decaying in time as a consequence of our results. Indeed, consider the following SDE:

\[
dX_t^{0,x} = -X_t^{0,x} \left(1 + \mathbb{E}[\exp(-|X_t^{0,x}|^2)]\right) dt + dB_t, \quad t \geq 0, \quad X_0^{0,x} = x \in \mathbb{R};
\]

and the quadratic test function.

Due to constant, non–degenerate diffusion, this example satisfies Assumption 3.2.2. Assumptions 3.3.2, 3.3.1 with \( p_1 = 0 \) and property (B.1.1) with any \( p \) and \( M_2 = 0, M_1 = 1, M_0 = 0 \) hold as well because of constant diffusion, linearly growing drift and boundedness of \( x \mapsto e^{-|x|^2} \) and its derivatives. Moreover, since it has additive noise and the drift is Lipschitz and monotonic due to the boundedness of \( x \mapsto e^{-|x|^2} \) and its derivatives, we know that it satisfies Assumptions 3.2.1 and 3.2.3. By Corollary 3.4.1, we obtain the following decay in time for some \( c > 0 \):

\[
|\partial_x V(t, x)| \leq C(1 + |x|^2) e^{-cs}, \quad \forall (s, x) \in (1, \infty) \times \mathbb{R}^d.
\]
Chapter 4

Weak error of subsampled approximation of McKean-Vlasov SDEs

The value of a function of solution to an SDE with non-linear drift in the sense of McKean is usually approximated using the empirical mean of the associated $N$–particle system. Recently a more computationally efficient alternative for sampling has been proposed: one can use another associated, randomly subsampled approximating system instead of the usual $N$–particle one. Although there is plenty of evidence of a performance improvement with this approach, quantitative results are missing in the literature. In this chapter we study first the problem from a continuous-time perspective and afterwards we present an uniform in time bound for the order of weak error in the Euler scheme approximation for both the $N$–particle system and the subsampled one. Extra detail is introduced by additional dependence of an equally subsampled external measure into the coefficient, which is of outstanding relevance for applications to 1–layer neural networks.

4.1 Introduction

Many fields such as physics, biology or sociology deal with modelling of collective behaviour resulting from a huge number of interrelated and exchangeable individuals. Mathematically, this study takes the shape of existence and uniqueness theorems for the solution to the weakly interacting particle system but also for the solution to its mean-field limit. Usually, the object of study of this field is precisely the limit, which is proved in some cases to satisfy a McKean-Vlasov SDE. This phenomenon takes the name of propagation of chaos. Now, this is only one side of the story. The other one is to study the almost never explicitly solvable McKean-Vlasov SDE by approximating its solution by sampling from the associated $N$-particle system. The word “approximating” already indicates that we must recover the original McKean-Vlasov SDE in the limit as $N \to \infty$.

There are two approaches for studying the error: strong and weak error (see Section 1.5.1) and they complement each other. Namely, while the strong error is given by the second moment of the difference in between the actual value and its approximation, the weak error studies the bias of such an approximation. In particular, for particle system approximations: the strong and the weak error receive the names of strong and weak propagation of chaos.
whereby the bias might vanish quicker. Indeed, simulating by the fact that since the strong error is driven by the statistical error, the weak error (or moments. This was later generalized in [57] to McKean–Vlasov SDEs with additive noise and

Euler scheme. This is done for McKean-Vlasov SDEs without external measure dependence

2.17] provides a power decomposition of the weak error's order in terms of the size of the

On Wasserstein spaces proved weak error order

The study of the weak error is not new: between others, [66] in the context of one dimensional viscous scalar conservation laws and afterwards [81] and [38] in the context of calculus on Wasserstein spaces proved weak error order \( O(1/N) \). More specifically, [38, Theorem 2.17] provides a power decomposition of the weak error's order in terms of the size of the system for general test functions on measure space which have enough derivatives in the sense of Lions (see his lectures notes at the Collège du France [82], which were redacted by Cardaliaguet). More recently, [81, Theorem 3.6] provides optimal rate propagation of chaos for test functions on measure spaces, for which one needs uniform ellipticity and bounded and globally Hölder continuous coefficients and test functions which have two bounded, Hölder continuous derivatives in both space and measure components. In terms of discretization results, [57] already proved weak error order \( O(1/N + h) \), where \( h \) the step size of the Euler scheme. This is done for McKean-Vlasov SDEs without external measure dependence and whose non–linearity in the sense of McKean is provided through dependence on the moments. This was later generalized in [57] to McKean–Vlasov SDEs with additive noise and possibly discontinuous drift, meant to accommodate rank–based models (the coefficients only depend on the rank of the \( i^{th} \) particle in the system).

The study of the weak error it is motivated from the numerical efficiency perspective by the fact that since the strong error is driven by the statistical error, the weak error (or bias) might vanish quicker. Indeed, simulating \( \sqrt{N} \) independent copies of the system with
\(\sqrt{N}\) particles leads to the same order of error for bias \(O(N^{-1/4} \times N^{-1/4})\) as for strong error \(O(\sqrt{N})\). However, when simulating a system with \(N\) particles, the bias \((= O(N^{-1}))\) is smaller than the strong error \((= O(\sqrt{N}))\). Both approaches have a resulting global error of \(O(\sqrt{N})\), although the former approach is less expensive and has a bigger bias than the latter as soon as the computational cost of the particle system grows more than linearly with the number of particles. The reason why not considering the alternative with smaller bias and the same global error is that sampling from a \(N\)-particle system is costly.

It has been noticed that using the whole empirical measure for the approximating particle system might not be the most efficient option since sampling from the full \(N\)-particle system with \(N'\)-data points input from the external measure has a cost, meaning computational cost per each time step, of \(N'^2 + NN'\). Consequently, the need to improve the efficiency when sampling has increasingly gained relevance in recent studies. Taking inspiration from the literature, both Markov Chain Monte Carlo Method and Stochastic Gradient Descent use random batches in order to reduce the size of the sample at each iteration step. The authors of [83] propose reducing the computational cost to an order \((S + S')N\) for some \(S \ll N\), \(S' \ll N'\) by subsampling random batches for each measure of \(S\) and \(S'\) particles and data points independently at each fixed time on the discretization grid, and considering only the empirical measure of each subsample when solving the above system of SDEs (4.1.2). In other words, at each fixed time, one picks randomly one subset of particles of size \(S\) and another subset of data points from the external measure of size \(S'\) and those are the only ones considered in each of the empirical measures input in the measure components of the coefficients of the SDE satisfied by any particle in the system. While subsampling is a way to reduce the computational cost of sampling from a particle system, it has been shown that it increases the strong error. A full computational complexity analysis and how the strong error is affected by subsampling is analysed in [83, See Section 2 and Corrolary 3.1] for dynamics whose drift is obtained from a confining potential plus a binary interaction kernel. Namely they obtain uniformly strong error order \(O(\sqrt{h} + N^{-1/2+\epsilon})\) for any \(\epsilon > 0\), where \(h\) is the step size. Hence, in this scenario the global error from subsampling would be \(O(\sqrt{h} + N^{-1/2+\epsilon})\), while the computational cost per time step is \(SN\). In this chapter we analyse this problem from the weak error's point of view, meaning that we study the effect of subsampling on the weak error order.

As a side note, we mention that the present work studies only the subsampling with replacement. It is hinted in [83] that not allowing replacement might reduce variance in some cases but the theoretical analysis is analogous and the cost reduction is the same.

More proof of the relevance of the subsampling method is presented in [84], where the authors explain the benefits of a possible application of subsampling to neural networks known by the name of “dropout”. Backed up by numerous data bases tests such as MNIST and TIMIT, the authors explain how overfitting can be avoided by randomly dropping out half of the data when training a neural network and so forcing neurons to rely on the whole population behaviour rather than on specific units. Additionally, “dropout” can be viewed as a very efficient model averaging technique compared to alternatives such as Bayesian model averaging or “bagging”. Finally, benefits of a specific type of dropout, called “Naive Bayes” is explained to show significant improvement with respect to the regular logistic classification when there is little training data. A more technical study of “dropout” can be found on [85], where the authors obtain some estimates of the regularizing and averaging properties which afterwards are corroborated with results from simulations.
In the general McKean–Vlasov case we illustrate first in continuous time (meaning that the subsampling occurs in the continuous time scheme), in absence of an external measure and assuming that the drift has a linear dependence on the measure, that we recover order \(O\left(\frac{1}{N}\right)\) for the weak error of a subsampled approximating particle system (see Theorem 4.2.3). In particular, the size of the subsample does not affect the order of the error, making subsampling a valuable technique since it reduces cost (hypothetically once we discretize) but does not worsen accuracy. However, if the dependence is not linear, the subsampling technique returns a weak error order which deteriorates with the batch size \(S\) (see Example 4.2.6). Even more proof of how much of a special case this independence of the size of the batch is, is presented in Example 4.2.5 where the strong error is of at least order \(O\left(\frac{1}{S}\right)\).

For a fully implementable algorithm however, one must discretise the time interval (into size \(h\) partitions). To that end, we assume a priori the existence of \(R\), some function representing a bound on the variance of the subsamples of sizes \(S\) and \(S'\) for the law of the state process and the external measure respectively. Our main result, Theorem 4.4.2, presents the following weak error order for the Euler scheme subsampled approximating particle system (although only after making some assumptions on the shape of the non-linearity allowed in (4.1.1)): \(O\left(\frac{1}{N} + h + h^{1/2}R(S, S')\right)\). In particular, while the weak error from subsampling would be \(O\left(\sqrt{h} + N^{-1/2+\epsilon}\right)\), the computational cost per time step is \(SN\). A couple of remarks are in order here: first, a new phenomenon arises— the size of the batches and the amount of time steps considered in the discrete scheme are expected to balance each other. Qualitatively, since in the discrete scenario we only sample at finite time points, the reduction in cost by sampling using this random batch method adds up to being considerable. Secondly and more importantly, the estimates we obtain are uniform in time, the uniformity rooting from the decaying in time bounds for the first, second and third derivatives of the solution to the associated backward Kolmogorov equation which we present in Proposition 4.3.10. This field of study has interest on its own and only recently it has been proved for the torus in [60] when the diffusion is constant and the drift belongs to some specific family such as the ones with small dependence on the measure. The problem boils down to the fact that uniform in time propagation of chaos might fail even in the scenario that a unique, globally attractive invariant measure to the Fokker-Planck equation associated to the McKean–Vlasov SDE exists (see [86]).

Finally, let us motivate the additional dependence on an external measure. The key is that our formulation can now accommodate models such as the ones studied in [41], with applications to deep learning. Namely, [41] presents an alternative way of approaching a non-convex minimization problem by lifting the problem to the measure space (using Lions derivative [35, Ch. 5], [82]) and consequently transforming it into a problem of convex minimization of a function on measure spaces. In Example 4.1.3 it is shown that such problems fall under our assumptions and therefore we justify, with a rigorous analysis of the weak error, the commonly spread practice of subsampling in the field of machine learning.

This chapter is organised as follows: in Section 2 we introduce rigorously the subsampling technique and analyse, with the help of various examples, its impact on the weak and strong error’s order. This allows us to single out the case when the drift is linear in measure which, despite it’s specific behaviour in terms of the size of the batch, can be used to illustrate the main ideas for obtaining a more general result in Section 4. Section 3 is a parenthesis in which we derive the necessary derivatives estimates for the solution to the backward Kolmogorov equation. We do this on the basis of a monotonicity assumption and employ derivative
processes. Since these estimates are of interest on their own, we present them in a higher
generality than what is needed for this chapter. In Section 4 we prove uniform in time weak
error estimates for the Euler discretization of the subsampled approximating particle system
associated to (4.1.1). Finally in Appendices C.4 and C.5 we include generalization to uniform
estimates of the weak error induced by the discretization of the McKean–Vlasov SDE and the
one induced by the discretization of the fully coupled $N$–particle system, respectively.

\subsection{PDE theory for weak error analysis}

Our analysis of the weak error is based on the corresponding PDE theory. It is well known (see
e.g. \cite[Theorem 7.6]{1} or \cite{5}) that given sufficiently regular initial data $\phi = \phi(x)$ and coefficients
$b = b(x)$ and $\sigma(x) = \sigma(x)$, the solution to the PDE:

$$
\begin{align*}
\partial_t U(t,x) - \frac{1}{2} \text{tr}(\sigma(x) \partial_x^2 U(t,x) \sigma^*(x))) &= \partial_x U(t,x) b(x), \quad (t,x) \in [0,\infty) \times \mathbb{R}^d; \\
U(0,x) &= \phi(x), \quad x \in \mathbb{R}^d,
\end{align*}
$$

is $U(s,x) = \mathbb{E}[\phi(X^x_s)]$, where for any $t \geq 0, x \in \mathbb{R}^d$ fixed,

$$
dx^{t,x}_s = b(X^{t,x}_s) \, ds + \sigma(X^{t,x}_s) \, dB_s, \; \forall s \in [t,\infty); \; X^{t,x}_t = x.
$$

Consider now a McKean–Vlasov SDE for given $t \geq 0, b: \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d$ and
$
\sigma: \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}^d \times \mathbb{R}^d :$

$$
dx^{t,\xi}_s = b(X^{t,\xi}_s, \mathcal{L}(X^{t,\xi}_s)) \, ds + \sigma(X^{t,\xi}_s, \mathcal{L}(X^{t,\xi}_s)) \, dB_s, \; \forall s \in [t,\infty); \; X^{t,\xi}_t = \xi.
$$

Assume that the coefficients $b$ and $\sigma$ are sufficiently nice so that (4.1.4) has a unique (in
distribution) solution $(X^{t,\xi}_s)_{s \in [t,\infty)}$ (e.g. \cite[Theorem 2.10, Corollary 3.4 and Theorem 4.1]{39},
\cite{88}). Since the uniqueness is in distribution, any other initial data with the same law
will lead to the same solution, in the equation above is enough to specify $\mu := \mathcal{L}(\xi)$.

Notice that, for each $\omega \in \Omega$, $X^{0,x}_{r+s}(\omega) \neq X^{s,x^0_r(\omega)}_{r+s}(\omega)$ for any $r > 0$, $x \in \mathbb{R}^d$, i.e. the flow
property doesn’t hold in $\mathbb{R}^d$, and so the process $(X^{\xi}_s)_{s \in [0,\infty)}$ is not a Markov process on $\mathbb{R}^d$.
From this it is clear that we would not expect $V(t,\xi) := \mathbb{E}[\phi(X^{\xi}_t)]$ to satisfy a PDE of the
form (4.1.3). However, it can be shown that $V$ will satisfy a PDE on $[0,\infty) \times L^2(\Omega)$, which
requires analysis of PDEs on Hilbert spaces. In this chapter, we study an alternative for
finding appropriate regularity estimates.

For the purpose of this chapter, it is possible to bypass PDEs on Hilbert spaces as follows.
Consider the following SDE:

$$
Y^{t,x,\xi}_s = x + \int_t^s b(Y^{t,x,\xi}_r, \mathcal{L}(Y^{t,x,\xi}_r)) \, dr + \int_t^s \sigma(Y^{t,x,\xi}_r, \mathcal{L}(Y^{t,x,\xi}_r)) \, dB_r, \; \forall s \in [t,\infty), \; x \in \mathbb{R}^d.
$$

Notice that $\mathcal{L}(X^{t,\xi}_s)$ is given by solving (4.1.4) and it is an input to the SDE (4.1.5). In particular,
this is a classical SDE, for which we assume there exists an unique strong solution. This implies
that $Y^{0,x,\xi}_{r+s}(\omega) = Y^{s,x^0_r(\omega)}_{r+s}(\omega)$ and this process is a Markov process on $\mathbb{R}^d$.
Consequently, for any fixed $\xi$, the classical Feynman–Kac Theorem holds for $V(t, x; \xi) := \mathbb{E}[\phi(Y^{\xi}_t)]$ (see

86
Assumption 4.1.1 (Regularity and growth coefficients)

\[ \xi \]
\[ \partial, V(t, x; \xi) - \frac{1}{2} \text{tr}(\sigma(x, \mathcal{L}(X_t^0, \xi))) \partial_x^2 V(t, x; \xi) \sigma^*(x, \mathcal{L}(X_t^0, \xi))) = (\partial_x V(t, x; \xi), b(x, \mathcal{L}(X_t^0, \xi))) ; \]
\[ V(0, x; \xi) = \phi(x). \]

(4.1.6)

One consequence of the fact that (4.1.5) has a unique strong solution is that for any \( \xi \) fixed, \( Y^{x, \xi} \) is a Borel-measurable map of \( x \) and \( B := (B_t)_{t \in [0, \infty)} \) is a well defined stochastic process. Moreover, because the solution to (4.1.4) is unique we have that \( Y^{x, \xi}_t = X^t \) for all \( t \in [0, \infty) \) (see Appendix C.2). This, together with the independence of \( \xi \) and \( W \), implies that for \( \mu = \mathcal{L}(\xi) \),

\[ \mathcal{V}(t, \xi) = \mathbb{E}[\phi(X_t^{x, \xi})] = \mathbb{E}[\phi(Y_t^{x, \xi})] = \int_{\mathbb{R}^d} \mathbb{E}[\phi(Y_t^{x, \xi})] \mu(dx) = \int_{\mathbb{R}^d} V(t, x; \xi) \mu(dx). \]

Based on this relation, one can borrow PDE results for the equation (4.1.6) and conclude results for \( \mathcal{V}(t, \xi) \). In the rest of the chapter we indiscriminately work with PDE results for the equation (4.1.6) in the sense that although as a consequence of the Chain Rule, for any \( i = 1, \ldots, d \)

\[ \partial_{x_i} V(t, x) = \partial_{x_i} V(t, x; \xi)_{\xi=x} + \partial_{\xi_i} V(t, x; \xi)_{\xi=x}, \quad \forall (t, x) \in [0, \infty) \times \mathbb{R}^d, \]

in our computations we will only be needing \( \partial_{x_i} V(t, x; \xi)_{\xi=x} \). This is because all our bounds are derived from increments of the form \( \mathbb{E}[\phi(Y_t^{x, \xi}) - \phi(Y_t^{x', \xi})] \) for some \( x, x' \in \mathbb{R}^d \). This explanation also motivates our abuse of notation when using \( \partial_{x_i} V(t, x) \) to denote \( \partial_{x_i} V(t, x; \xi)_{\xi=x} \) and \( \partial_{\xi_i}^a V(t, x) \) to denote \( \partial_{\xi_i}^a V(t, x; \xi)_{\xi=x} \) for any \( |a| = 0, 1, 2, 3, 4 \).

### 4.1.2 Assumptions

On the probability space \((\Omega, \mathcal{F}, \mathbb{P})\) we have defined a \( d \)-dimensional Brownian motion \( B \), an independent random variable \( \xi \in L^{p_*}(\mathbb{R}^d) \) (\( p_* \) imposed by Assumption 4.1.3) and \( (B_t^{i})_{t \in [1, \ldots, N]} \) independent \( d \)-dimensional Brownian motions together with \( (\xi_t^i)_{t \in [1, \ldots, N]} \in L^{p_*}(\mathbb{R}^d) \) i.i.d copies of \( \xi \).

**Assumption 4.1.1** (Regularity and growth coefficients). In terms of regularity we ask for the following:

1. The functions \( b \in C^{4,0,0}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d)) \) and \( \sigma \in C^{4,0}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)) \), where the continuity in the measure components is in the Wasserstein–1 distance. Moreover, for each \( n = 1, 2, 3, 4 \), there exist \( I^n \in \mathbb{N} \) such that \( \partial_x^n b(\cdot, \mu, \nu) \in B_{1,n}(\mathbb{R}^d) \) and \( \partial_x \sigma(\cdot, \nu) \in C^3_b(\mathbb{R}^d) \), all of it uniformly in \( \mu \in \mathcal{P}_2(\mathbb{R}^d) \) and \( \nu \in \mathcal{P}_2(\mathbb{R}^d) \).

   Moreover, \( \sup_{s \geq 0} (|b(s, 0)| + |\Sigma(s, 0)|) < \infty \).

2. There exist \( C_{Lip} > 0 \) and \( p \in \mathbb{N} \), \( p \geq 2 \), such that for all \( (x, \mu, \nu), (x', \mu', \nu') \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d) \),

\[ |b(x, \mu, \nu) - b(x', \mu', \nu')| + |\sigma(x, \nu) - \sigma(x', \nu')| \]
\[ \leq C_{\text{Lip}} \left( (1 + |x|^{p/2} + |x'|^{p/2})|x - x'| + \mathcal{W}_1(\mu, \mu') + \mathcal{W}_1(v, v') \right). \]

**Assumption 4.1.2** (Regularity and growth of test functions). The function \( \phi : \mathbb{R}^d \to \mathbb{R} \) is Lipschitz, four times continuously differentiable (i.e \( \phi \in C^4(\mathbb{R}^d) \)). Moreover, for \( n = 0, 1, 2, 3, 4 \) there exists \( \mathbb{N} \ni I_n \) such that \( \partial^n \phi \in B_{I_n}(\mathbb{R}^d) \).

**Assumption 4.1.3** (Monotonicity). Let \( p^* \geq \max\{4p + 5, 16 \} \max \{I_i\} + 1, 32 \max \{I_i\} + 1, 65 \} \). For any \( v \in \mathcal{P}_2(\mathbb{R}^d) \), there exist some \( M_1(v) > 2M_0(v) > 0 \) such that for any \( (x, \mu), (x_0, \mu_0) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \),

\[
\langle x - x_0, b(x, \mu, v) - b(x_0, \mu_0, v) \rangle + \frac{p^* - 1}{2} ||\sigma(x, v) - \sigma(x_0, v)||_{HS}^2 \leq -M_1(v)|x - x_0|^2 + M_0(v) \int_{\mathbb{R}^d} |y - y_0|^2 d\mu(y) d\mu(y_0). \tag{4.1.7}
\]

Notice that given the finite dimensional setup, all the matrix norms are equivalent. And since we do not care about the change in the constants, when is more convenient we use the trace norm instead.

**Assumption 4.1.4** (Subsampling bias and variance). Given \( N, N' \in \mathbb{N} \) and some \( 0 < S' \leq N' \), let \( \Theta^S \) and \( \Theta^{S'} \) be uniform random variables from \( \{1, \ldots, N\} \) onto the sets \( \{1, \ldots, S\} \) and \( \{1, \ldots, S'\} \) respectively. Fixed \( v \in \mathcal{P}_2(\mathbb{R}^d) \), we assume that for all \( (N, S), (N', S') \), \( \mu \in \mathcal{P}_2(\mathbb{R}^d) \) there exist \( \mathcal{R}(S, S'), \mathcal{R}(S, S') \in \mathbb{R} \) (small) satisfying the following small bias and variance estimator conditions.

For any i.i.d random variables \( \{X^i\}_{i=1,...,N} \) such that \( X^1 \sim \mu \) and another i.i.d random variables \( \{Y^i\}_{i=1,...,N'} \) such that \( Y^1 \sim v \) and finally for all \( x \in \mathbb{R}^d \),

\[
\left| E \left[ b \left( x, \frac{1}{S} \sum_{j=1}^{S} \delta_{X^{\Theta^S(j)}}, \frac{1}{S'} \sum_{j=1}^{S'} \delta_{Y^{\Theta^{S'}(j)}} \right) \right] - b(x, \mu, v) \right| + \left| E \left[ \sigma \left( x, \frac{1}{S} \sum_{j=1}^{S} \delta_{Y^{\Theta^{S'}(j)}} \right) \right] - \sigma(x, v) \right| \leq \mathcal{R}(S, S'),
\]

and:

\[
E \left[ \left| b \left( x, \frac{1}{S} \sum_{j=1}^{S} \delta_{X^{\Theta^S(j)}}, \frac{1}{S'} \sum_{j=1}^{S'} \delta_{Y^{\Theta^{S'}(j)}} \right) - b(x, \mu, v) \right|^2 + \left| \sigma \left( x, \frac{1}{S} \sum_{j=1}^{S} \delta_{Y^{\Theta^{S'}(j)}} \right) - \sigma(x, v) \right|^2 \right] \leq \mathcal{R}(S, S')(1 + |x|^2).
\]

Next we mention a few remarks related to this set of assumptions and provide an example for which Assumption 4.1.4, possibly the most surprising one, holds:

1. We next show that Assumption 4.1.1 implies a certain growth condition over the coefficients. Indeed, a particular implication of Assumption 4.1.1.2 is that for all \( (x, \mu, v) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d) \),

\[
|b(x, \mu, v) - b(0, \delta_0, \delta_0)| \leq C_{\text{Lip}}(1 + |x|^{p/2})|x| + C_{\text{Lip}} \int_{\mathbb{R}^d} |x|^2 \mu(dx) + C_{\text{Lip}} \int_{\mathbb{R}^d} |x|^2 v(dx).
\]

88
Next by the triangle inequality, if $|b(0, \delta_0, \delta_0)| + |\sigma(0, \delta_0)| < \infty$, then there exists some constant $C > 0$ such that the following is satisfied for all $(x, \mu, \nu) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d)$:

$$|b(x, \mu, \nu)| \leq C \left( (1 + |x|^{p/2}) |x| + \int_{\mathbb{R}^d} |x|^2 \mu(dx) + \int_{\mathbb{R}^d} |x|^2 \nu(dx) \right).$$  \hfill (4.1.8)

Moreover, because in Assumption 4.1.1.1 we asked that once fixed $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, $\partial_x \phi(\cdot, \nu) \in C_b(\mathbb{R}^d)$, we conclude that there exists $C > 0$ such that

$$|\sigma(x, \nu)| \leq C (1 + |x|), \quad \forall x \in \mathbb{R}^d. \hfill (4.1.9)$$

2. $\mathbb{R}^d \times \mathbb{R}^{d \times N} \times \mathcal{P}_2(\mathbb{R}^d) \ni (x, y, \nu) \mapsto b(x, y, \nu) := b\left(x, \frac{1}{N} \sum_{i=1}^N \delta_{y_i}, \nu\right)$, is Lipschitz in the $y$ component, uniformly in $(x, \nu) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$. Indeed, for any $y, y' \in \mathbb{R}^{d \times N}$, as a consequence of the Lipschitz condition in Assumption 4.1.1.2, the duality formula of the 1–Wasserstein distance (which allows us to obtain homogeneity) and its behaviour when applied to empirical measures,

$$|b(x, y, \nu) - b(x, y', \nu)|$$

$$= \left| b\left(x, \frac{1}{N} \sum_{i=1}^N \delta_{y_i}, \nu\right) - b\left(x, \frac{1}{N} \sum_{i=1}^N \delta_{y_i'}, \nu\right) \right| \leq C_{Lip} \mathcal{W}_1\left( \frac{1}{N} \sum_{i=1}^N \delta_{y_i}, \frac{1}{N} \sum_{i=1}^N \delta_{y_i'} \right)$$

$$= C_{Lip} \frac{1}{N^2} \mathcal{W}_1\left( \sum_{i=1}^N \delta_{y_i}, \sum_{i=1}^N \delta_{y_i'} \right) \leq C_{Lip} \frac{1}{N} \sum_{i=1}^N \mathcal{W}_1(\delta_{y_i}, \delta_{y_i'}) \leq C_{Lip} \frac{1}{N} \sum_{i=1}^N |y_i - y_i'|$$

$$\leq C |y - y'|.$$  \hfill (4.1.10)

3. Although we assume throughout the chapter that $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, it can be observed throughout the chapter that the results still hold for measures in $\mathcal{P}_2(E)$, where $E$ is any complete metric space. Another immediate generalization is considering $n$–dimensional Brownian Motion (with $n$ not necessarily equal to $d$).

4. Assumptions 4.1.1–4.1.3 gives existence and uniqueness of solution to the considered SDE (4.1.1) (see [76, 87, 88]). Moreover, Assumption 4.1.3 implies a coercivity condition (see (4.3.5)) which we prove that implies finite moments of order up to $p^* - 1$ (see Appendix C.2).

5. Notice that Assumption 4.1.4 assumes control over the bias and the variance, conditions which can be tested numerically for each particular problem at hand. To show that this is not an empty assumption, we provide Example 4.1.5 below. An example is also presented as [89, Example 2.15], where they also give a hint on how to reduce variance of the error by possibly defining subsampling without replacement. In terms of the bias, i.e. the first part of Assumption 4.1.4, we notice that a particular case is the unbiased estimator scenario. This, we prove in (4.2.4), is the case of a linear (in the measure component) drift. In this scenario one enjoys some cancellations, which are noticeable in the bound of the weak error.
Example 4.1.5. Consider the drift $b(x, \mu, \nu) := \mathbb{E}[\alpha(x, X)]$, where $X \sim \mu$. Then, for $(X^j)_{j=1, \ldots, N}$ i.i.d random variables distributed as $\mu$, its projection through the empirical mean is defined as:

$$b\left(x, \frac{1}{N} \sum_{j=1}^{N} \delta_{X^j}\right) = \frac{1}{N} \sum_{j=1}^{N} \alpha(x, X^j).$$

Notice that as proved previously in identity (4.2.4), given the linear dependence of the drift in the measure component, we have that the subsampled drift is an unbiased estimator of the one considering all the particles, i.e. for all $1 \leq S \leq N$:

$$\mathbb{E}\left[ b\left(x, \frac{1}{N} \sum_{j=1}^{N} \delta_{X^j}\right) - b\left(x, \frac{1}{S} \sum_{k=1}^{S} \delta_{X^{\Theta^{S}(k)}}\right) \right] = \mathbb{E}\left[ \frac{1}{N} \sum_{j=1}^{N} \alpha(x, X^j) - \frac{1}{N} \sum_{k=1}^{N} \frac{1}{S} \sum_{j=1}^{S} \alpha(x, X^k) \right] = 0.$$  

(4.1.10)

Then, since $(X^j)_{j=1, \ldots, N}$ are i.i.d and due to (4.1.10) (i.e $\frac{1}{N} \sum_{j=1}^{N} \alpha(x, X^j) - \frac{1}{S} \sum_{k=1}^{S} \alpha(x, X^{\Theta^{S}(k)})$ is a centred expression) and if moreover we assume $\sup_{X \in \mathbb{L}^2(\mathbb{R}^d)} \alpha(x, X) = C(1 + |x|^2)$ for all $x \in \mathbb{R}^d$, we have:

\[
\begin{align*}
\mathbb{E}\left[ b\left(x, \frac{1}{N} \sum_{j=1}^{N} \delta_{X^j}\right) - b\left(x, \frac{1}{S} \sum_{k=1}^{S} \delta_{X^{\Theta^{S}(k)}}\right) \right]^2 & \\
= \mathbb{E}\left[ \left( \frac{1}{N} \sum_{j=1}^{N} \alpha(x, X^j) - \frac{1}{S} \sum_{k=1}^{S} \alpha(x, X^{\Theta^{S}(k)}) \right)^2 \right] \\
= \frac{1}{S^2} \mathbb{E}\left[ \left( \sum_{k=1}^{S} \left( \alpha(x, X^{\Theta^{S}(k)}) - \frac{1}{N} \sum_{j=1}^{N} \alpha(x, X^j) \right)^2 \right) \right] \\
= \frac{1}{S^2} \sum_{k=1}^{S} \mathbb{E}\left[ \left( \alpha(x, X^{\Theta^{S}(k)}) - \frac{1}{N} \sum_{j=1}^{N} \alpha(x, X^j) \right)^2 \right] \\
= \frac{1}{S^2} \sum_{k=1}^{S} \sum_{j'=1}^{N} \mathbb{E}\left[ \left( \alpha(x, X^{j'}) - \frac{1}{N} \sum_{j=1}^{N} \alpha(x, X^j) \right)^2 \right] \\
= \frac{1}{S^2} \sum_{j'=1}^{N} \mathbb{E}\left[ \left( \alpha(x, X^{j'}) - \frac{1}{N} \sum_{j=1}^{N} \alpha(x, X^j) \right)^2 \right] \\
& \leq \frac{1}{S N} \sum_{j'=1}^{N} \mathbb{E}\left[ \left( \frac{1}{N} \sum_{j=1}^{N} \alpha(x, X^j) \right)^2 \right] \\
& \leq \frac{1}{S N} \sum_{j'=1}^{N} \mathbb{E}\left[ \left( \frac{1}{N} \sum_{j=1}^{N} \alpha(x, X^j) \right)^2 \right] \\
& \leq \frac{1}{S} C(1 + |x|^2).
\end{align*}
\]
The extra condition on \( \alpha \) is achieved for example by \( \alpha(x, y) = -x(e^{-y^2} + 1), x, y \in \mathbb{R} \). The fact that the function is smooth and for any \( y \in \mathbb{R}, e^{-y^2} + 1 \in [1, 2] \) makes it easy to check that Assumptions 4.1.1, 4.1.3, 4.1.4 are satisfied for examples with such a drift and constant diffusion.

Ideally, one would prove empirically that for the problem at hand, \( R(S) \) is of order \( \frac{1}{S^2} \). This would imply that indeed, subsampling does not affect the order of the weak error if \( S \) is taken equal to \( \frac{p}{N} \). This is the case of the simplest example: additive noise and drift such that \( b \left( \frac{1}{N} \sum_{j=1}^{N} \delta X_j \right) = -\frac{1}{N} \sum_{j=1}^{N} X_j \).

4.1.3 Motivational example

The practical motivation for obtaining uniform in time orders of the weak error when the underlying dynamics depends on an external measure is partially supported by its application to 1-layer neural networks. In the following example it is shown that such problems fall under the set of assumptions proposed in Section 4.1.2 and therefore we provide an analytical justification to the common practice of subsampling in machine learning.

In a nutshell, [41] presents an alternative way of approaching a non-convex minimization problem. Namely, the problem is lifted to the measure space where it becomes a problem of convex minimization of functions of measure. Namely, consider the following minimization problem \( \min_{\mu} \{ F(\mu) + \frac{\sigma^2}{2} H(\mu) \} \), where \( H \) is the relative entropy with respect to a Gibbs measure of which the density is proportional to \( e^{-U(x)} \) and which is added as a regularizer. Then, fixed accordingly \( \sigma > 0 \) and under appropriate conditions over the functions, the flow of measures solutions to the associated Langevin dynamics:

\[
dX_t = - \left( D_m F(\mu_t, X_t) + \frac{\sigma^2}{2} \nabla U(X_t) \right) dt + \sigma dB_t, \quad t \geq 0; \quad \mu_t = \mathcal{L}(X_t),
\]

converges to an invariant measure \( \mu^* \) such that \( \min_{\mu} \{ F(\mu) + \frac{\sigma^2}{2} H(\mu) \} = F(\mu^*) + \frac{\sigma^2}{2} H(\mu^*) \).

In the particular case that one is training a one layer neural network, the functional \( F \) takes the following shape:

\[
F(\mu) := \int_{\mathbb{R}^d} \Phi(y - \mathbb{E}^\mu[\hat{\phi}(X; z)]) v(dz, dy),
\]

where \( X \sim \mu \) represents the parameters of the network, \( \mathbb{E}^\mu[\hat{\phi}(X; z)] \) is the output of the trained network with input \( z \), \( \Phi: \mathbb{R} \to \mathbb{R} \) is a given potential function and \( v \) is the data distribution. Notice that although the dependence on the external measure \( v \) is not explicitly stated in \( b(x, \mu) := -D_m F(\mu, x) - \frac{\sigma^2}{2} \nabla U(x) \), it does inherit its dependence on \( v \) from \( F \).

Conceptually, the functional \( F \) represents, for \( \Phi(x) := x^2 \), the mean square error of the neural network trained with the given data. Now, in this particular case, the intrinsic measures derivatives is explicitly known:

\[
\frac{\delta F}{\delta m}(\mu, x) = - \int_{\mathbb{R}^d \times \mathbb{R}} \Phi'(y - \mathbb{E}^\mu)[\hat{\phi}(X, z)] \hat{\phi}(x, z) v(dz, dy),
\]

and
\[ D_mF(\mu, x) = -\int_{\mathbb{R}^d \times \mathbb{R}} \Phi'(y - \mathbb{E}^\mu[\hat{\phi}(X, z)]) \nabla \hat{\phi}(x, z) \nu(dz, dy). \]

Now back to the general problem (4.1.11), under the following assumptions, [41, Theorem 2.10] claims existence of an invariant measure minimising the function \( F \) and to which the flow of measures \( (\mu_t)_{t \geq 0} \) converges.

(A) \( F \in C^1(\mathcal{P}(\mathbb{R}^d)) \) (see [35, Chapter 5] or [38, Section 2.1] for an introduction to calculus on measure spaces) is convex and bounded from below.

(B) \( D_mF : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \), representing the intrinsic derivative of \( F \) satisfies:

1. \( D_mF \) is bounded and Lipschitz continuous in both its components, i.e. there exists \( C_F > 0 \) such that for all \( x, x' \in \mathbb{R}^d; m, m' \in \mathcal{P}(\mathbb{R}^d) \),
   \[ |D_mF(m, x) - D_mF(m', x')| \leq C_F(|x - x'| + \mathcal{W}(m, m')). \]

2. \( D_mF(m, \cdot) \in C^\infty(\mathbb{R}^d) \) for all \( m \in \mathcal{P}(\mathbb{R}^d) \).

3. \( \nabla D_mF : \mathcal{P}(\mathbb{R}^d) \times \mathbb{R}^d \rightarrow \mathbb{R}^d \) is jointly continuous.

4. Moreover, there exists \( C_F > 0 \) such that for all \( m \in \mathcal{P}(\mathbb{R}^d) \),
   \[ |D_mF(m, 0)| \leq C_F \left(1 + \int_{\mathbb{R}^d} |y|dm(y)\right). \]

In [41, Assumption 3.2], sufficient conditions are stated for \( \Phi \) and \( \hat{\phi} \) under which this Assumption (B) holds for (4.1.12).

(C) \( U : \mathbb{R}^d \rightarrow \mathbb{R} \) belongs to \( C^\infty(\mathbb{R}^d) \), \( \partial_x U \) is Lipschitz continuous and there exist \( C_U > 0, C'_U \in \mathbb{R} \) such that
   \[ \langle \partial_x U(x), x \rangle \geq C_U|x|^2 + C'_U, \forall x \in \mathbb{R}^d; \]
   and
   \[ \langle \partial_x U(x) - \partial_x U(x'), x - x' \rangle \geq C_U|x - x'|^2, \forall x, x' \in \mathbb{R}^d. \]

Let us check that this set of assumptions imply ours, that way allowing us to apply to this example all the theory developed in this chapter. First notice that for any \( m \in \mathcal{P}(\mathbb{R}^d) \) fixed, the three times continuous differentiability for \( D_mF(m, \cdot) \) follows from Assumption (B.2). Moreover, the joint continuity is implied by the joint continuity of its gradient, stated in Assumption (B.3). The fact that the drift and the diffusion are jointly Lipschitz in all its variables follows from \( \sigma \) being constant and from Assumptions (B.1) and (C) above. Altogether Assumption 4.1.1 holds. Moreover, from Assumptions (B.1), (B.4) and (C) we obtain using Jensen’s inequality with \( \varepsilon = C_F^{-1/4} \):

\[
\begin{align*}
\langle x - x_0, -D_mF(\mu, x) - \frac{\sigma^2}{2} \partial_x (U(x)) + D_mF(\mu_0, x_0) + \frac{\sigma^2}{2} \partial_x (U(x_0)) \rangle \\
\leq |\langle x - x_0, D_mF(\mu, x) - D_mF(\mu_0, x_0) \rangle| - \frac{\sigma^2}{2} |\langle x - x_0, \partial_x (U(x)) - \partial_x (U(x_0)) \rangle| \\
\leq \varepsilon^2 \frac{|x - x_0|^2}{2} + \frac{1}{2\varepsilon^2} |D_mF(\mu, x) - D_mF(\mu_0, x_0)|^2 - \frac{\sigma^2}{2} C_U|x - x_0|^2
\end{align*}
\]

92
\[
\begin{align*}
&= \left( \frac{1}{2 \epsilon^2} |D_m F(\mu, x) - D_m F(\mu_0, x_0)|^2 - \frac{\sigma^2}{2} C_U \right) - \left( \frac{\sigma^2}{2} C_U - \frac{\epsilon^2}{2} \right) |x - x_0|^2 \\
&= \left( \frac{\sqrt{C_F}}{2} \mathcal{W}_t(\mu, \mu_0) \right) - \left( \frac{\sigma^2}{2} C_U - \frac{\sqrt{C_F}}{2} \right) |x - x_0|^2.
\end{align*}
\]

Finally, notice that at the expense of picking a big enough diffusion constant \( \sigma \) or when the constants \( C_F, C_U \) are well balanced, we do get that \( M_1 := \frac{\sigma^2}{2} C_U - \frac{\sqrt{C_F}}{2} > 0 \), which means that Assumption 4.1.3 is also satisfied. Finally, Assumption 4.1.2 is checked computationally on a case by case basis.

## 4.2 Subsampling discussion

The aim for this section is to make the idea of subsampling (with replacement) mathematically rigorous. More explicitly, we define the size \( S \in \{1, \ldots, N\} \subset \mathbb{N} \) subsampling from a \( N \)-particle system by introducing an extra random process \( (\Theta^S_t)_{t \geq 0} \), taking values in the subset \( \{1, \ldots, S\} \subset \mathbb{N} \), which is uniformly distributed in \( \{1, \ldots, N\} \subset \mathbb{N} \).

**Definition 4.2.1** (Random batches). For \( t \geq 0, 0 < S < N \in \mathbb{N} \),

\[ \Theta^S_t : \tilde{\Omega} \to \{f : \{1, \ldots, S\} \to \{1, \ldots, N\}\} \]

is a family of random maps defined on a new probability space \( (\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}) \), such that:

- **They are uniformly distributed random variables** in the sense that for any \( t \geq 0 \) and \( \forall j \in \{1, \ldots, S\} \) and \( \forall j' \in \{1, \ldots, N\} \), \( \tilde{\mathbb{P}}(\Theta^S_t(j) = j') = \frac{1}{N} \);

- **Batches are independent over time**, i.e. for all \( t \geq 0 \) and \( s \in [0, t] \),

\[ \tilde{\mathbb{P}}(\Theta^S_t(j) = j' | (\Theta^S_s : s \leq t)) = \frac{1}{N}, \quad \forall j = 1, \ldots, S, \forall j' = 1, \ldots, N. \]

The subsampled approximating system for (4.1.1) on \( (\Omega \times \tilde{\Omega}) \times \mathcal{F} \otimes \tilde{\mathcal{F}} \otimes \mathcal{F}, \mathbb{P} \otimes \tilde{\mathbb{P}} \otimes \mathbb{P} \) (these subspaces have their own expectation operators and should be denoted differently, but when there is no confusion, we will denote all of them by \( \mathbb{E} \) is):

\[
\begin{cases}
    dZ^{i,N,S}_t &= b(Z^{i,N,S}_t, \mu^{Z,N,\Theta^S_t}_t, \nu^N, \phi^S_t) \, dt + \sigma(Z^{i,N,S}_t, \nu^N, \phi^S_t) \, dB^i_t, \quad i = 1, \ldots, N; \\
    Z^{i,N,S}_0 &= \xi^i,
\end{cases}
\]

where \( \mu^{Z,N,\Theta^S_t}_t := \frac{1}{S} \sum_{j=1}^{S} \delta_{Z^{i,N,S}_t(j)} \), while the full empirical measure will be denoted by

\[ \mu^{Z,N}_t := \frac{1}{N} \sum_{j=1}^{N} \delta_{Z^{i,N}_t(j)} \].

Also, \( \nu^N, \phi^S_t := \frac{1}{S} \sum_{j=1}^{S} \delta_{y^{i,N,\phi^S_t(j)}} \), where the family of \( N' \)-sized data \( \{y^i\}_{i=1, \ldots, N'} \) are sampled from the distribution \( \nu \). Notice that for a more intuitive notation, we should denote the particle system by \( \{(Z^{i,N,S,N',S'}_t)_{t \geq 0}\}_{i=1, \ldots, N} \). However, in order to avoid cumbersome notation, we do not include the dependence of the particle system on \( S', N' \) when these are fixed.
Of course, working with a subset of particles instead of the whole set of them induces extra error. The question to be answered at this point is how much the weak error worsens by subsampling and whether it is of some use to subsample this way, since the final goal is still to balance computational cost and weak error accuracy. Moreover for this algorithm to hold any practical relevance, we must consider its time discretization and we do so in Section 4.4. However, aiming to motivate the study of the subsampling error, we present first a continuous time analysis where we consider the additional simplification that the subsampled drift and diffusion are “unbiased” and we do not have dependence on an external measure. Our main result in this section (Theorem 4.2.3) states that the order of weak error of the $S$–sized subsampled system is $O(1/N)$.

In the further sections we will consider instead a more general framework where we do have dependence on an external measure, the dependence on the law of the solution is less restrictive and moreover an Euler scheme is used to discretize the approximating particle system. In this scenario, the uniform in time order $\tilde{O}(1/N + h + h^{1/2} \mathcal{R}(S, S') + \tilde{\mathcal{R}}(S, S'))$ for the weak error is presented in Theorem 4.4.2.

To conclude, we want to underline a few ideas:

- Theorems 4.2.3 and 4.4.2 generalize the result presented by Bencheikh and Jourdain in [57], where we include subsampling and dependence on an external measure.

- The variance of the subsampling, in the of Assumption 4.1.4 and represented here by $\mathcal{R}$, is proved in some cases (see [38]) to be only $O\left(\frac{1}{S} + \frac{1}{S'}\right)$. This implies that, in these cases, subsampling doesn’t save any time since it increases too much the error. It can however be used with storage–space saving goals in mind. Moreover, in Section 4.2.1, we prove that in the case of a linear dependence on the measure subsampling reduces computational cost at no expense on the weak error’s order. Which means that subsampling saves time and space in a linear scenario.

- After discretization, the size of the time step $h$ and the sizes of the random batches balance each other. Intuitively, as time steps get smaller, we shuffle more frequently the particles considered in the random batches and the bias induced by the subsampling bothers less. Therefore, since the bias is represented by a non–increasing function of the size of the batches ($\mathcal{R}$), the size of the batches can be smaller as well and so the computational cost of simulating can be balanced.

- From this uniform in time bound of the weak error’s order of the discrete scheme, together with the results in the chapter [41], we conclude that the same order of weak error is achieved when an invariant measure is being approximated by the empirical law of a sub-sampled approximating particle system.

### 4.2.1 Subsampling in continuous time

For this very first motivational case, we consider the following McKean-Vlasov SDE with linear dependence on the measure and additive noise:

$$dX_t^\xi = E[\alpha(X_t^\xi)] dt + dB_t, \quad t \in [0, \infty); \quad X_0^\xi = \xi,$$  \hspace{1cm} (4.2.2)
where $\alpha : \mathbb{R}^d \to \mathbb{R}^d$ is a Lipschitz function. Let it be approximated, once fixed $S, N \in \mathbb{N}$ such that $0 < S \leq N$, by:

$$
\begin{cases}
    dZ_t^{i,N,S} = \frac{1}{S} \sum_{j=1}^{S} \alpha(Z_t^{\Theta_t^S(j),N,S}) dt + dB_t^i, & t \in [0, \infty), i = 1, \ldots, N;
    
    Z_0^{i,N,S} = \xi^i,
\end{cases}
$$

where for each $t \in [0, \infty)$, the random batch $\Theta_t^S$ is given by Definition 4.2.1.

We first present the lemma at the base of our PDE approach.

**Lemma 4.2.2.** Fix $T > 0$, $S, N \in \mathbb{N}$ such that $0 < S \leq N$ and let Assumptions 4.1.1 and 4.1.4 hold. Consider for given $\xi$, a square–integrable random variable, the processes $X^\xi$ and $Z^{i,N,S}$ with dynamics given by (4.2.2) and (4.2.3) respectively. Finally, let $V$ be the solution to PDE (4.1.6) for functions $\phi$ satisfying Assumption 4.1.2. Then, there exists $C(T) > 0$ independent of $N, S$ such that

$$
\left| \mathbb{E} \left[ \int_0^T \frac{1}{N} \sum_{k=1}^{N} \left\langle \partial_x V(s, Z^{k,N,S}_{T-s}), \sum_{j=1}^{N} \frac{1}{N} \alpha(Z_t^{j,N,S}) - \mathbb{E}[\alpha(X_t)] \right\rangle ds \right] \right| \leq \frac{C(T)}{N}.
$$

Notice the dependence of the bound on the size of the time interval. This can be avoided using more sophisticated bounds (requiring monotonicity Assumption 4.1.3) and we will do so for the discrete case study. However, considering the illustrative purpose of this section, we would like to focus more on the PDE–based spirit the proof and the dependence of the weak error with respect to $S$ and $N$ rather than $T$.

**Proof of Lemma 4.2.2.** Since $\xi$ is fixed and square–integrable, let us simplify the notation as $X = X^\xi$ for the remainder of the proof.

Adding and subtracting the same expression and applying Cauchy–Schwarz inequality, we obtain:

$$
\left| \mathbb{E} \left[ \int_0^T \frac{1}{N} \sum_{k=1}^{N} \left\langle \partial_x V(s, Z^{k,N,S}_{T-s}), \sum_{j=1}^{N} \frac{1}{N} \alpha(Z_t^{j,N,S}) - \mathbb{E}[\alpha(X_t)] \right\rangle ds \right] \right| 
= \left| \mathbb{E} \left[ \int_0^T \left\langle \frac{1}{N} \sum_{k=1}^{N} \partial_x V(s, Z^{k,N,S}_{T-s}) - \mathbb{E}[\partial_x V(s, X_{T-s})], \sum_{j=1}^{N} \frac{1}{N} \alpha(Z_t^{j,N,S}) - \mathbb{E}[\alpha(X_t)] \right\rangle ds \right] \right|
+ \mathbb{E} \left[ \int_0^T \mathbb{E}[\partial_x V(s, X_{T-s})], \sum_{j=1}^{N} \frac{1}{N} \alpha(Z_t^{j,N,S}) - \mathbb{E}[\alpha(X_t)] \right\rangle ds \right] \right| 
\leq \left( \mathbb{E} \left[ \int_0^T \left\langle \partial_x V(s, Z^{k,N,S}_{T-s}) - \mathbb{E}[\partial_x V(s, X_{T-s})] \right\rangle^2 ds \right] \right)^{1/2}
+ \left( \mathbb{E} \left[ \int_0^T \sum_{j=1}^{N} \frac{1}{N} \alpha(Z_t^{j,N,S}) - \mathbb{E}[\alpha(X_t)] \right\rangle^2 ds \right) \right)^{1/2}
\left| \mathbb{E} \left[ \int_0^T \left\langle \partial_x V(s, X_{T-s})], \sum_{j=1}^{N} \frac{1}{N} \alpha(Z_t^{j,N,S}) - \mathbb{E}[\alpha(X_t)] \right\rangle ds \right] \right|.
$$

Now, by using Fubini and the strong propagation of chaos result – Theorem C.1.1 – we
can bound the first summand in the previous expression as follows:
\[
\left( \int_0^T \mathbb{E} \left[ \frac{1}{N} \sum_{k=1}^N \partial_x V(s, Z_{T-s}^{k,N,S}) - \mathbb{E}[\partial_x V(s, X_{T-s})] \right]^2 ds \right)^{1/2} \\
\quad \times \left( \int_0^T \mathbb{E} \left[ \sum_{j=1}^N \frac{1}{N} \alpha(Z_s^{j,N,S}) - \mathbb{E}[\alpha(X_s)] \right]^2 ds \right)^{1/2} \\
\leq C(T) \left( \frac{1}{N^{1/2}} \right) \left( \frac{1}{N^{1/2}} \right).
\]

Next we make a parenthesis to note the following un–bias property. Given the family of random processes \(\{Z_t^{i,N,S}, i = 1, \ldots, N\}\) consider the sigma–algebra generated by it, i.e. \(\mathcal{F}_t^Z := \sigma((Z_s^{j,N,S} : s \in [0, t]; i = 1, \ldots, N))\) and denote the conditional expectation with respect to it as \(\mathbb{E}^{Z,t}[\cdot] := \mathbb{E}[\cdot | \mathcal{F}_t^Z]\). We will be using the key observation that, by definition of the random variables \(\Theta\),
\[
\mathbb{E}^{Z,t}\left[ \frac{1}{S} \sum_{j=1}^S \delta_{Z_t^{j,N,S}} \right] = \frac{1}{S} \sum_{j=1}^S \sum_{f=1}^N \sum_{j=1}^N \delta_{Z_t^{f,N,S}} = \frac{1}{N} \sum_{j=1}^N \sum_{j=1}^N \delta_{Z_t^{f,N,S}}.
\]
In particular, for any given function \(\alpha\),
\[
\mathbb{E}^{Z,t}\left[ \alpha(Z_t^{N,j,N,S}) \right] = \frac{1}{N} \sum_{j=1}^N \alpha(Z_t^{j,N,S}). \tag{4.2.4}
\]

In other words, in the case that the drift is linear, the subsampled drift \(\frac{1}{S} \sum_{j=1}^S \alpha(Z_t^{j,N,S})\) is an unbiased estimator for the drift evaluated in the whole empirical measure of the subsampled particle system, i.e. \(\frac{1}{N} \sum_{j=1}^N \alpha(Z_t^{j,N,S})\).

Then, by (4.2.4),
\[
\left| \mathbb{E}\left[ \int_0^T \left( \mathbb{E}[DV(s, X_{T-s})] \sum_{j=1}^N \frac{1}{N} \alpha(Z_s^{j,N,S}) - \mathbb{E}[\alpha(X_s)] \right) ds \right] \right| \\
= \left| \int_0^T \mathbb{E}[DV(s, X_{T-s})] \mathbb{E}\left[ \sum_{j=1}^N \frac{1}{N} \alpha(Z_s^{j,N,S}) - \mathbb{E}[\alpha(X_s)] \right] ds \right| = 0.
\]

Putting all together we conclude the bound \(\frac{C(T)}{N}\).

Uniform in time bounds can be achieved with some extra manipulations by avoiding the use of strong propagation of chaos (which is of interest on its own) and appealing instead to the decay in time of the gradient of \(V\).

**Theorem 4.2.3.** Let \(S, N\) be two natural numbers such that \(0 < S \leq N\). Under the assumptions in Lemma 4.2.2, let us approximate (4.2.2) (given a square–integrable random variable \(\xi\)) by (4.2.3). Then, for all test functions \(\phi\) satisfying Assumption 4.1.2 and all \(T > 0\), there exists \(C = C(T) > 0\) independent of \(N, S\) such that
\[
\left| \mathbb{E}[\phi(X_T^{\xi})] - \frac{1}{N} \sum_{j=1}^N \mathbb{E}[\phi(Z_T^{j,N,S})] \right| \leq \frac{C}{N}.
\]
Proof. Since $\xi$ is fixed and square-integrable, let us simplify the notation as $X = X^\xi$ for the remainder of the proof.

A key observation is that due to the fact that the random variables are identically distributed, we can reformulate the weak error as

$$\left| E[\phi(X_T)] - \frac{1}{N} \sum_{j=1}^{N} E[\phi(Z_{T}^{j,N,S})] \right| = \left| E[\phi(X_T)] - E[\phi(Z_{T}^{1,N,S})] \right|. $$

Next, recall $V(t, x) = E[\phi(X_t^x)]$ and so in particular $E[\phi(X_T)] = E[V(T, X_0)] = E[V(T, Z_{T}^{1,N,S})]$. Moreover, from the initial condition of the PDE (4.1.6), we have that $E[\phi(Z_{T}^{1,N,S})] = E[V(0, Z_{T}^{1,N,S})]$. Hence it all leads to a further reformulation of the weak error’s expression

$$E[\phi(X_T)] - E[\phi(Z_{T}^{1,N,S})] = E[V(T, Z_{T}^{1,N,S})] - E[V(0, Z_{T}^{1,N,S})].$$

Notice that this is the right shape where Itô’s formula, paired with the PDE (4.1.6), can be applied to obtain:

$$E[V(T, Z_{T}^{1,N,S}) - V(0, Z_{T}^{1,N,S})]$$

$$= E \left[ \int_0^T \partial_s V(s, Z_{T-s}^{1,N,S}) - \left\langle \partial_x V(s, Z_{T-s}^{1,N,S}), \sum_{j=1}^{S} \frac{1}{S} \alpha(Z_{s}^{j,N,S}) \right\rangle - \frac{1}{2} \mbox{tr} (\partial_x^2 V(s, Z_{T-s}^{1,N,S})) ds \right]$$

$$= E \left[ \int_0^T \left\langle \partial_x V(s, Z_{T-s}^{1,N,S}), E[\alpha(X_s)] - \sum_{j=1}^{S} \frac{1}{S} \alpha(Z_{s}^{j,N,S}) \right\rangle ds \right].$$

Let us use now the tower property and the fact the particles are independent and identically distributed to conclude:

$$\left| E[\phi(X_T)] - \frac{1}{N} \sum_{j=1}^{N} E[\phi(Z_{T}^{j,N,S})] \right|$$

$$= E \left[ E^Z \left[ \int_0^T \frac{1}{N} \sum_{k=1}^{N} \left\langle \partial_x V(s, Z_{T-s}^{k,N,S}), \sum_{j=1}^{S} \frac{1}{S} \alpha(Z_{s}^{j,N,S}) - E[\alpha(X_s)] \right\rangle ds \right] \right].$$

Next, we applying Fubini’s Theorem and take the conditional expectation inside the integral. Moreover, due to the compatibility condition being satisfied, the sigma algebra on which we are conditioning can be changed from $\sigma((Z_{T}^{i,N,S} : i = 1, \ldots, N; 0 \leq r \leq T))$ to $\sigma((Z_{T-s}^{i,N,S} : i = 1, \ldots, N; 0 \leq r \leq s))$. Indeed, this compatibility condition is nothing else than appropriate measurability: by the Definition 4.2.1, the random shuffling is independent of the initial data and the Brownian Motion, which as a consequence imply that the integrand

$$\frac{1}{N} \sum_{k=1}^{N} \left\langle \partial_x V(s, Z_{T-s}^{k,N,S}), \sum_{j=1}^{S} \frac{1}{S} \alpha(Z_{s}^{j,N,S}) - \sum_{j=1}^{N} \frac{1}{N} \alpha(Z_{s}^{j,N,S}) \right\rangle$$

is independent on the future of the filtration generated by $Z^{N,S}$ (the sigma-algebra $\sigma((Z_{r-s}^{i,N,S} : i = 1, \ldots, N; s < r \leq T))$ given the present of the same filtration (i.e $\scr{F}^N = \sigma((Z_{r-s}^{i,N,S} : i = 1, \ldots, N; 0 \leq r \leq s))$). For more technical details on compatibility see [35, Chapter 1. Vol. II].
Hence, from (4.2.4), after taking conditional expectation inside the time integral,

$$
|E\phi(X_T)| - \frac{1}{N} \sum_{j=1}^{N} E[\phi(Z^{j,N,S}_T)]
$$

$$
= E\left[ E^{Z,T}\left[ \int_{0}^{T} \frac{1}{N} \sum_{k=1}^{N} \left\langle \partial_x V(s, Z^{k,N,S}_{T-s}), \sum_{j=1}^{S} \alpha(Z^{\Theta_{j}^{(j)},N}_{s}) - E[\alpha(X_s)] \right\rangle ds \right] \right]
$$

$$
= E\left[ \int_{0}^{T} \frac{1}{N} \sum_{k=1}^{N} E^{Z,T}\left[ \left\langle \partial_x V(s, Z^{k,N,S}_{T-s}), \sum_{j=1}^{S} \alpha(Z^{\Theta_{j}^{(j)},N,S}_{s}) - E[\alpha(X_s)] \right\rangle \right] ds \right]
$$

$$
= E\left[ \int_{0}^{T} \frac{1}{N} \sum_{k=1}^{N} \left\langle \partial_x V(s, Z^{k,N,S}_{T-s}), E^{Z,s}\left[ \sum_{j=1}^{S} \alpha(Z^{\Theta_{j}^{(j)},N,S}_{s}) - E[\alpha(X_s)] \right] \right\rangle ds \right].
$$

Finally, from the remark leading to identity (4.2.4), we arrive at the expression bounded by Lemma 4.2.2 and we conclude as follows:

$$
= E\left[ \int_{0}^{T} \left\langle \frac{1}{N} \sum_{k=1}^{N} \partial_x V(s, Z^{k,N,S}_{T-s}), \sum_{j=1}^{S} \alpha(Z^{j,N,S}_{s}) - E[\alpha(X_s)] \right\rangle ds \right]
$$

$$
\leq \frac{C(T)}{N}.
$$

\[ \square \]

**Remark 4.2.4.** Let us discuss the independence on the size of the batches. First, since we are considering shuffling in continuous time, even if the “particles” are subdivided in confined batches at each fixed time, the shuffling occurs frequently enough over the interval [0, T] for all “particles” to be able to encounter each other frequently enough for the effect of the subsampling to be unnoticeable. This technique is used in machine learning implementations as a dropout method for saving up memory and preventing co-adaptation [84, 90].

### 4.2.2 Dependence on the size of the batches

The goal of this subsection is to illustrate, using explicit examples, that the non-dependence of the error’s order on the size of the batch is a very fragile phenomenon.

We first show an example where the order of strong error, even in the linear measure dependence scenario, is affected heavily by the size of the random batches. The example is presented by making a parallel comparison between the sub-sampled and the usual/non-subsampled approximating systems.

**Example 4.2.5.** We consider the following McKean–Vlasov SDE, a particular case of that considered in Section 4.2.1 where \( \alpha \) is taken to be the identity:

$$
\frac{dX_t^\zeta}{\xi} = -E[X_t^\zeta] \, dt + dB_t, \quad t \in [0, \infty); \quad X_0^\zeta = \zeta. \tag{4.2.5}
$$

Here, based on linearity, we can compute explicitly the expectation of the solution by solving

$$
\frac{dE[X_t^\zeta]}{\xi} = -E[X_t^\zeta] \, dt \quad \text{to obtain} \quad E[X_t^\zeta] = E[\zeta] e^{-t}.
$$

And when substituted in the drift above:

$$
\frac{dX_t^\zeta}{\xi} = -E[\zeta] e^{-t} \, dt + dB_t,
$$

which finally gives us the explicit solution

$$
X_t^\zeta = \zeta + (e^{-t} - 1)E[\zeta] + B_t, \forall t \geq 0.
$$
On one hand, the usual associated approximating particle system for (4.2.5) is the following one:

\[
\begin{cases}
    dY^i_{t,N} = -\frac{1}{N} \sum_{j=1}^{N} Y^j_{t,N} \, dt + dB^i_t, & i = 1, ..., N; \\
    Y^i_{0,N} = \xi^i.
\end{cases}
\]

Notice that after summing up over the index \(i\), we obtain the following SDE:

\[
d\left(\frac{1}{N} \sum_{i=1}^{N} Y^i_{t,N}\right) = -\left(\frac{1}{N} \sum_{i=1}^{N} Y^i_{t,N}\right) \, dt + \frac{1}{N} \sum_{i=1}^{N} dB^i_t,
\]

which consequently implies the identity: \(\frac{1}{N} \sum_{j=1}^{N} Y^j_{t,N} = e^{-t} \sum_{j=1}^{N} \xi^j + \frac{1}{N} \sum_{j=1}^{N} \int_0^t e^{-(t-s)} \, dB^j_s\).

On the other hand, consider the subsampled approximating system for (4.2.5) and some \(0 < S \leq N\):

\[
\begin{cases}
    dZ^i_{t,N,S} = -\frac{1}{S} \sum_{j=1}^{S} Z^{\Theta_j(i),N,S}_{t,N,S} \, dt + dB^i_t, & i = 1, ..., N; \\
    Z^i_{0,N,S} = \xi^i.
\end{cases}
\]

In particular, for some \(i \in \{1, ..., S\}, dZ^{\Theta_j(i),N,S}_t = -\frac{1}{S} \sum_{j=1}^{S} Z^{\Theta_j(i),N,S}_{t,N,S} \, dt + dB^{\Theta_j(i)}_t\). After summing up over \(i\) one obtains, in a similar fashion to the non-subsampled system, that

\[
d\left(\frac{1}{S} \sum_{j=1}^{S} Z^{\Theta_j(i),N,S}_t\right) = -\left(\frac{1}{S} \sum_{j=1}^{S} Z^{\Theta_j(i),N,S}_t\right) \, dt + \frac{1}{S} \sum_{i=1}^{S} dB^{\Theta_j(i)}_t
\]

Hence,

\[
\frac{1}{S} \sum_{j=1}^{S} Z^{\Theta_j(i),N,S}_t = \frac{e^{-t}}{S} \sum_{j=1}^{S} \xi^{\Theta_0(j)} + \frac{1}{S} \sum_{j=1}^{S} \int_0^t e^{-(t-s)} \, dB^{\Theta_{j-1}}_s.
\]  

(4.2.6)

An important observation is the symmetry of the setup (i.i.d random variables and definition of \(\Theta\)), which has as a consequence the fact that for all \(j \in \{1, ..., N\},
\[
\mathbb{E}[X^j_t] = \mathbb{E}[Y^j_{t,N}] = \mathbb{E}[Z^j_{t,N,S}], \quad \forall t \geq 0.
\]

In this context, let us consider the mean square error for the non-subsampled system, which by independence of \(\xi\) and \(B\) is equal to:

\[
\mathbb{E}\left[\left|\mathbb{E}[X^\xi_t] - \frac{1}{N} \sum_{j=1}^{N} Y^j_{t,N}\right|^2\right] = \text{Var}\left(\frac{1}{N} \sum_{j=1}^{N} Y^j_{t,N}\right)
\]

\[
= \text{Var}\left(\frac{e^{-t}}{N} \sum_{j=1}^{N} \xi^j\right) + \text{Var}\left(\frac{1}{N} \sum_{j=1}^{N} \int_0^t e^{-(t-s)} \, dB^j_s\right)
\]

\[
= \frac{e^{-2t}}{N} \text{Var}(\xi) + \frac{1/2(1 - e^{-2t})}{N}.
\]

However, for the sub-sampled system, after using repeatedly the observation about the
absence of bias in the linear scenario (see (4.2.4)),

\[
\mathbb{E}\left[ \left| \mathbb{E}[X_t^j] - \frac{1}{N} \sum_{j=1}^{N} Z_{t}^{j,N,S} \right|^2 \right] = \mathbb{E}\left[ \left| \mathbb{E}[X_t^j] - \frac{1}{S} \sum_{j=1}^{S} Z_{t}^{\Theta_j(j),N,S} \right|^2 \right] + \mathbb{E}\left[ \left| \frac{1}{S} \sum_{j=1}^{S} Z_{t}^{\Theta_j(j),N,S} - \frac{1}{N} \sum_{j=1}^{N} Z_{t}^{j,N,S} \right|^2 \right]
\]

Now first by (4.2.6) and last by absence of bias in the linear scenario concluded in identity ((4.2.4)),

\[
\text{Var}\left( \frac{1}{S} \sum_{j=1}^{S} Z_{t}^{\Theta_j(j),N,S} \right) = \text{Var}\left( \frac{e^{-t}}{S} \sum_{j=1}^{S} \xi^{\Theta_j(j)} + \frac{1}{S} \sum_{j=1}^{S} \int_{0}^{t} e^{-(t-s)} dB_{s}^{\Theta_j(j)} \right)
\]

Finally, by the independence of the shuffles at different times (see Definition (4.2.1)),

\[
\mathbb{E}\left[ \left| \mathbb{E}[X_t^j] - \frac{1}{N} \sum_{j=1}^{N} Z_{t}^{j,N,S} \right|^2 \right] = \mathbb{E}\left[ \left| \mathbb{E}[X_t^j] - \frac{1}{S} \sum_{j=1}^{S} Z_{t}^{\Theta_j(j),N,S} \right|^2 \right] + \mathbb{E}\left[ \left| \frac{1}{S} \sum_{j=1}^{S} Z_{t}^{\Theta_j(j),N,S} - \frac{1}{N} \sum_{j=1}^{N} Z_{t}^{j,N,S} \right|^2 \right]
\]

The conclusion is that uniformly in time, although in the usual approximating system the order of strong error is \( \Theta\left( \frac{1}{N} \right) \), in the subsampled approximating system, the order increases (recall that \( S \ll N \)) to \( \Theta\left( \frac{1}{S} \right) \).

Next we present an explicit example in which the non-linearity of the drift breaks the order \( \Theta\left( \frac{1}{N} \right) \) of the weak error, concluded in Section 4.2.1 in a linear scenario. As a consequence, it shows that the subsampling method, although reducing computational cost in all cases, damages too much the weak error in some non-linear in measure drift scenarios to constitute any time reduction in the approximating problem.
Example 4.2.6. In (4.1.1), let us take diffusion coefficient $\sigma = 1$ and consider one of the simplest non-linear in measure drifts for 1-dimensional processes: given $\alpha : \mathbb{R} \to \mathbb{R}$ non-constant, bounded from above and below by two positive constants, define

$$b(x, \mu, \nu) = \left( \int_{\mathbb{R}^d} \alpha(x) d\mu(x) \right)^2 = \left( \mathbb{E}[\alpha(X)] \right)^2,$$ where $X \sim \mu$.

Consider then the same framework as in the previous sections, namely (4.1.1) approximated by (4.2.1) once fixed $S, N \in \mathbb{N}$ with $0 < S \leq N$. Then, just as in the proof of Theorem 4.2.3, the weak error’s expression can be rewritten as follows:

$$\left| \mathbb{E}\left[ \frac{1}{N} \sum_{j=1}^N \phi(Z_T^{j,N,S}) - \phi(X_T^L) \right] \right|$$

$$= \left| \mathbb{E}[V(T, Z_0^{1,N,S}) - V(0, X_T^L)] \right|$$

$$= \left| \mathbb{E} \left[ \int_0^T \partial_s V(s, Z_T^{1,N,S}(s)) \left( \sum_{i=1}^S \frac{1}{S} \alpha(Z_s^{\Theta_s^{(i)},N,S}) \right)^2 - \left( \sum_{i=1}^N \frac{1}{N} \alpha(Z_s^{i,N,S}) \right)^2 \right] ds \right|. \quad (4.2.7)$$

Moreover, due to the algebraic identity $a^2 - b^2 = (a - b)^2 - 2b(a - b), \forall a, b \in \mathbb{R}$, we conclude from above:

$$\left| \mathbb{E}\left[ \frac{1}{N} \sum_{j=1}^N \phi(Z_T^{j,N,S}) - \phi(X_T^L) \right] \right|$$

$$= \left| \mathbb{E} \left[ \int_0^T \partial_s V(s, Z_T^{1,N,S}(s)) \left( \sum_{i=1}^N \frac{1}{N} \alpha(Z_s^{i,N,S}) \right)^2 - \left( \sum_{i=1}^N \frac{1}{N} \alpha(Z_s^{i,N,S}) \right)^2 \right] ds \right|. \quad (4.2.7)$$

First notice that first because of the compatibility condition and second because of identity (4.2.4),

$$\mathbb{E} \left[ \int_0^T \partial_s V(s, Z_T^{1,N,S}(s)) \left( \sum_{i=1}^N \frac{1}{N} \alpha(Z_s^{i,N,S}) \right)^2 - \left( \sum_{i=1}^N \frac{1}{N} \alpha(Z_s^{i,N,S}) \right)^2 \right] ds$$

$$\mathbb{E} \left[ \int_0^T \partial_s V(s, Z_T^{1,N,S}(s)) \left( \sum_{i=1}^N \frac{1}{N} \alpha(Z_s^{i,N,S}) \right)^2 - \left( \sum_{i=1}^N \frac{1}{N} \alpha(Z_s^{i,N,S}) \right)^2 \right] ds$$

$$= 0.$$

Now, if the term

$$\mathbb{E} \left[ \int_0^T \partial_s V(s, Z_T^{1,N,S}(s)) \left( \sum_{i=1}^N \frac{1}{N} \alpha(Z_s^{\Theta_s^{(i)},N,S}) \right)^2 - \left( \sum_{i=1}^N \frac{1}{N} \alpha(Z_s^{i,N,S}) \right)^2 \right] ds,$$

were to be of a bigger order than $O\left( \frac{1}{N} \right)$, that would also mean order of weak error in (4.2.7) bigger that $O\left( \frac{1}{N} \right)$. Therefore, after recalling decay of $\partial_s V$ (see Section 4.3), it is enough to
show that there exists some positive constant $C$ such that

$$
\mathbb{E}\left[ Z^s \left[ \left( \sum_{j=1}^{S} \frac{1}{S} \alpha(Z_s^{j(i),N,S}) - \sum_{j=1}^{N} \frac{1}{N} \alpha(Z_s^{j,N,S}) \right)^2 \right] \right] \geq \frac{C}{S}.
$$

(4.2.8)

Notice that here is when the variance of the subsampling comes into play, given that the previous expression considers the second moment of the difference between the two measures $\mu^Z_{s,N,\theta^s}$ and $\mu^Z_{s,N,S}$. And from Example 4.2.5 we know that it is to be expected for the variance to be proportional to $1/S$.

In order to simplify notation, let us name $a := \sum_{j=1}^{N} \alpha(Z_s^{j,N,S})$ and $\bar{a} := \sum_{j=1}^{N} \frac{N}{S} \alpha(Z_s^{j,N,S})$. Then,

$$
\mathbb{E}\left[ Z^s \left[ \left( \sum_{j=1}^{S} \frac{1}{S} \alpha(Z_s^{j(i),N,S}) - \sum_{j=1}^{N} \frac{1}{N} \alpha(Z_s^{j,N,S}) \right)^2 \right] \right] = \frac{1}{N^2} \mathbb{E}\left[ Z^s \left[ \left( \sum_{j'=1}^{S} \frac{N}{S} \alpha(Z_s^{j'(i),N,S}) - \sum_{j=1}^{N} \alpha(Z_s^{j,N,S}) \right)^2 \right] \right] = \frac{1}{N^2} \mathbb{E}\left[ Z^s \left| a - \bar{a} \right|^2 \right].
$$

Moreover, notice that as a consequence of (4.2.4), we have the following centring identity for all $j' \in \{1, \ldots, N\}$:

$$
\mathbb{E}\left[ Z^s \left[ N\alpha(Z_s^{j'(i),N,S}) - \sum_{j=1}^{N} \alpha(Z_s^{j,N,S}) \right] = 0.
$$

This used repeatedly together with Definition 4.2.1 allow us to continue:

$$
\mathbb{E}\left[ Z^s \left| a - \bar{a} \right|^2 \right] = \mathbb{E}\left[ Z^s \left[ \frac{1}{S^2} \left( \sum_{j=1}^{S} N\alpha(Z_s^{j(i),N,S}) - S \sum_{j=1}^{N} \alpha(Z_s^{j,N,S}) \right)^2 \right] \right] = \frac{1}{S^2} \mathbb{E}\left[ Z^s \left[ \left( \sum_{j=1}^{S} N\alpha(Z_s^{j(i),N,S}) - S \sum_{j=1}^{N} \alpha(Z_s^{j,N,S}) \right)^2 \right] \right] = \frac{1}{S^2} \mathbb{E}\left[ Z^s \left[ \left( \sum_{j=1}^{S} N\alpha(Z_s^{j(i),N,S}) - S \sum_{j=1}^{N} \alpha(Z_s^{j,N,S}) \right)^2 \right] \right] = \frac{1}{S^2} \left( SN \sum_{j=1}^{N} \alpha(Z_s^{j(i),N,S})^2 - S a^2 \right) = \frac{1}{S} \left( N \sum_{j=1}^{N} \alpha(Z_s^{j(i),N,S})^2 - a^2 \right) = \frac{1}{S} \left( (N-1) \sum_{j=1}^{N} \alpha(Z_s^{j(i),N,S})^2 - \sum_{j_1,j_2=1}^{N} \alpha(Z_s^{j_1(i),N,S}) \alpha(Z_s^{j_2(i),N,S}) \right) = \frac{1}{S} \left( (N-1) \sum_{j=1}^{N} \alpha(Z_s^{j(i),N,S})^2 - \sum_{j_1=1}^{N-1} \sum_{j_2=j_1+1}^{N} \alpha(Z_s^{j_1(i),N,S}) \alpha(Z_s^{j_2(i),N,S}) \right)
$$
we work with a non-autonomous SDE rather than a McKean-Vlasov one, since we can take

\[ \frac{1}{S} \sum_{j=1}^{N} \left( \alpha(Z_{s}^{j,N,S}) - \alpha(Z_{s}^{j,N,S}) \right) \]

\[ = \frac{1}{S} \sum_{j=1}^{N} \left( \alpha(Z_{s}^{j,N,S}) - \alpha(Z_{s}^{j,N,S}) \right) \]

\[ = \frac{1}{S} \sum_{j,j_{1}=1}^{N} \left( \alpha(Z_{s}^{j,N,S}) - \alpha(Z_{s}^{j,N,S}) \right) \]

\[ = \frac{1}{S} \sum_{j,j_{1}=1}^{N} \left( \alpha(Z_{s}^{j,N,S}) - \alpha(Z_{s}^{j,N,S}) \right)1 \{ \alpha(Z_{s}^{j,N,S}) \geq 0 \} \]

\[ = \frac{1}{S} \sum_{j,j_{1}=1}^{N} \left( \alpha(Z_{s}^{j,N,S}) - \alpha(Z_{s}^{j,N,S}) \right)1 \{ \alpha(Z_{s}^{j,N,S}) \geq 0 \} \]

Finally, notice that under the assumption that \( \alpha \) is bounded from below and from above by

\[ \mathbb{E} [ \left| a^{\ast} - a \right| ] \geq \frac{N(N-1)C}{S}. \]

Hence, when returning to the original expression (4.2.8),

\[ \mathbb{E} [ \left| a^{\ast} - a \right| ] \geq \frac{N(N-1)C}{S}. \]

And indeed we have shown that the order of the weak error is at least \( \Theta \left( \frac{1}{S} \right) \).

### 4.3 Derivatives decay for solutions to the backward Kolmogorov equation

In this section we present explicit decaying (in the time component) bounds of part of the

\[ \phi(X_{s}^{0,x}) \]

\[ \rightarrow \mathcal{V}(s,x) = \mathbb{E} [ \phi(X_{s}^{0,x}) ] \]

\[ (X_{s}^{0,x}) \]

\[ \text{is the process satisfying (4.1.4) with deterministic initial condition } x \in \mathbb{R}^{d} \text{ and } \phi, \text{a function satisfying Assumption 4.1.2.} \]

We do this by bounding the moments of the corresponding derivative processes. However, we work with a non-autonomous SDE rather than a McKean-Vlasov one, since we can take advantage of the uniqueness of solution to (4.1.4) guaranteed by Assumptions 4.1.1 and 4.1.3 (see [76, Theorem 6.1]) (in this section we generalized to \( \sigma : \mathbb{R}^{d} \times \mathcal{F}_{2}(\mathbb{R}^{d}) \rightarrow \mathbb{R}^{d} \times \mathbb{R}^{d} \) continuously differentiable in the measure component, with uniformly bounded derivative and satisfying the rest of Assumption 4.1.1 uniformly in \( \mu \)). Indeed, let us continue the discussion we started in Section 4.1.1. The unique solution to (4.1.4) will give us, for any \( p^{*}-\)
integrable random variable \( \xi \), the value of \( \mathcal{L}(X_{s}^{0,\xi}) \) which can be plugged into the coefficients \( b \) and \( \sigma \) in order to obtain the following non-autonomous coefficients:

\[
\beta(s, x; \xi) = b(x, \mathcal{L}(X_{s}^{0,\xi})) \quad \text{and} \quad \Sigma(s, x; \xi) = \sigma(x, \mathcal{L}(X_{s}^{0,\xi})).
\]

Based on this, from (4.1.1) and for fixed \( \tau \in [0, \infty) \) and \( \psi \), a \( p^* \)-integrable random variable, we obtain the following related SDE on \( [\tau, \infty) \):

\[
dY^{\tau, \psi; \xi}_{s} = \beta(s, Y^{\tau, \psi; \xi}_{s}; \xi) d s + \Sigma(s, Y^{\tau, \psi; \xi}_{s}; \xi) d B_{s}; \quad Y^{\tau, \psi; \xi}_{\tau} = \psi.
\] (4.3.1)

We can now define the auxiliary function \( V(t, \psi; \xi) = \mathbb{E}[\phi(Y_{s}^{0,\xi})] \), but in order to be free to work with this function instead, it remains to check that the bounds we obtain remain valid at some level for the function of interest \( V(s, x) = \mathbb{E}[\phi(X_{s}^{0,\xi})] \). A first relation is collected in Lemma C.2.1, which states that \( X^{\psi}_{s} = Y_{s}^{0,\xi} \) a.s. for all \( s, \xi \), meaning that \( V(s, x) = V(s, x; x) \) for all \( s, x \). However notice that derivatives–wise we only have that for all \( s, x \in [0, \infty) \times \mathbb{R}^{d} \),

\[
\partial_x V(s, x) = \partial_{\psi} V(s, \psi; \xi)|_{\psi=\xi=x} + \partial_\xi V(s, \psi; \xi)|_{\psi=\xi=x}.
\]

Similarly, for higher order derivatives we need the crossed derivatives as well. The good news is that for purposes such as weak error estimates, it is enough to obtain bounds for just \( \partial_\psi V(s, x; x) \) (as we will show below; see also [57]). With respect to the notation, let us clarify once more that because we are interested in this particular derivative, we are going to use the following indiscriminately: \( Y_{s}^{0,x;x} = Y_{s}^{x} \) and

\[
\partial_x V(s, x) = \partial_{\psi} V(s, \psi; \xi)|_{\psi=\xi=x}.
\]

Moreover, in line with this notation simplification, for \( n = 0, 2, 3, 4 \) we abuse the notation for the derivatives and define:

\[
\partial^n_x \beta(s, x) = \partial^n_{\psi} \beta(s, \psi; \xi)|_{\psi=\xi=x} \quad \text{and} \quad \partial^n_x \Sigma(s, x) = \partial^n_{\psi} \Sigma(s, \psi; \xi)|_{\psi=\xi=x}.
\]

In Section 4.1.2, we introduced the regularity and monotonicity conditions. We state them next in the more general context which includes measure–dependent diffusion and a more relaxed monotonicity condition.

**Assumption 4.3.1 (Regularity and growth coefficients).** In terms of regularity we ask for the following:

1. Fixed \( v \in \mathcal{P}_2(\mathbb{R}^{d}) \), the functions \( b(\cdot, \cdot, v) \in \mathcal{C}^{4,0}(\mathbb{R}^{d} \times \mathcal{P}_2(\mathbb{R}^{d})) \) and \( \sigma(\cdot, \cdot, v) \in \mathcal{C}^{4,0}(\mathbb{R}^{d} \times \mathcal{P}_2(\mathbb{R}^{d})) \), where the continuity in the measure component is in the 1–Wasserstein distance. Moreover, for each \( n = 1, 2, 3, 4 \), there exist \( l_n \in \mathbb{N} \) such that \( \partial^n_x b(\cdot, \mu, v) \in B_{l_n}(\mathbb{R}^{d}) \) and \( \partial_x \sigma(\cdot, \mu, v) \in C_{b}^{3}(\mathbb{R}^{d}) \), all of it uniformly in \( \mu \in \mathcal{P}_2(\mathbb{R}^{d}) \).

2. Fixed \( v \in \mathcal{P}_2(\mathbb{R}^{d}) \), there exist \( C_{Lip} > 0 \) and \( p \in \mathbb{N} \), \( p \geq 2 \), such that for all \( (x, \mu), (x', \mu') \in \mathbb{R}^{d} \times \mathcal{P}_2(\mathbb{R}^{d}) \),

\[
|b(x, \mu, v) - b(x', \mu', v)| + |\sigma(x, \mu, v) - \sigma(x', \mu', v)| \leq C_{Lip} \left((1+|x|^{p/2}+|x'|^{p/2})|x-x'| + \mathcal{W}_1(\mu, \mu')\right).
\]
**Assumption 4.3.2** (Monotonicity). For any \( v \in \mathcal{P}_2(\mathbb{R}^d) \), there exist some \( M_1 > 2M_0 > 0 \), \( M \in \mathbb{R} \) and \( p^* \geq \max \{65,p + 3,32i + 1\} \) such that for any \( (x, \mu), (x_0, \mu_0) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \),

\[
\langle x - x_0, b(x, \mu, v) - b(x_0, \mu_0, v) \rangle + \frac{p^* - 1}{2} \| \sigma(x, \mu, v) - \sigma(x_0, \mu_0, v) \|^2 \\
\leq -M_1 \| x - x_0 \|^2 + M_0 \int_{\mathbb{R}^d} \| y - y_0 \|^2 d\mu(y) d\mu_0(y) + M.
\]

Now, for the coefficients of the auxiliary process (4.3.1), this property reads as follows.

**Lemma 4.3.3.** Suppose that Assumption 4.3.2 holds. Then, there exists \( \widetilde{M} \in \mathbb{R} \) such that for any \( s \geq 0 \) and for any \( x, x_0 \in \mathbb{R}^d \):

\[
\langle x - x_0, \beta(s,x) - \beta(s,x_0) \rangle + \frac{p^* - 1}{2} \| \Sigma(s,x) - \Sigma(s,x_0) \|^2 \\
= \langle x - x_0, b(x, \mathcal{L}(X_s^{0,x_0}), v) - b(x, \mathcal{L}(X_s^{0,x_0}), v) \rangle + \frac{p^* - 1}{2} \| \sigma(x, \mathcal{L}(X_s^{0,x_x}), v) - \sigma(x_0, \mathcal{L}(X_s^{0,x_0}), v) \|^2 \\
\leq -M_1 \| x - x_0 \|^2 + M_0 \mathbb{E}[\| X_s^{0,x} - X_s^{0,x_0} \|^2] + M.
\]

Furthermore, from Lemma C.2.3, which provides explicit second moments bounds,

\[
\mathbb{E}[\| X_s^{0,x} - X_s^{0,x_0} \|^2] \leq e^{-2(M_1-M_0)s} \left( \| x - x_0 \|^2 - \frac{M}{M_1-M_0} \right) + \frac{M}{M_1-M_0}.
\]

Next, observe that \( 0 < \sup_{s \geq 0} M_0 e^{-2(M_1-M_0)s} = M_0 \) and \( \sup_{s \geq 0} \left( \frac{M}{M_1-M_0} \right) = \frac{M}{M_1-M_0} \).

After naming \( \tilde{M} = M + \frac{M_0 M}{(M_1-M_0)} \), all the above put together leads to:

\[
\langle x - x_0, \beta(s,x) - \beta(s,x_0) \rangle + \frac{p^* - 1}{2} \| \Sigma(s,x) - \Sigma(s,x_0) \|^2 \leq -(M_1 - M_0) \| x - x_0 \|^2 + \tilde{M}.
\]

Therefore, (4.3.2) is consequence of Assumption 4.3.2. \( \Box \)

### 4.3.1 Derivative processes

First we study the sensitivity with respect to the initial data and moments of the auxiliary process (4.3.1).

**Lemma 4.3.4.** Under Assumption 4.3.2, the process \( (Y_s^{X})_{s \geq 0} \) defined by (4.3.1) satisfies the following: there exists \( C > 0 \) such that for any \( 2 \leq m \leq p^* \) and \( x, x' \in \mathbb{R}^d \):

\[
\mathbb{E}[\| Y_s^{X} - Y_s^{X'} \|^m] \leq C (1 + \| x - x' \|^m e^{-m(M_1-M_0)s}), \quad \text{for all } s \geq 0.
\]
Proof. Let $\lambda$ be an arbitrary positive constant. If Itô’s formula is applied to the stochastic process $(e^{\lambda s} | Y_s^x - Y_s^{x'} |^m)_{s \geq 0}$, we obtain:

$$d\left(e^{\lambda s} | Y_s^x - Y_s^{x'} |^m\right) = e^{\lambda s}\left(\lambda | Y_s^x - Y_s^{x'} |^m + m | Y_s^x - Y_s^{x'} |^{m-2} \langle Y_s^x - Y_s^{x'}, (\beta(s, Y_s^x) - \beta(s, Y_s^{x'})) \rangle\right.
\left. + \frac{m(m-1)}{2} | Y_s^x - Y_s^{x'} |^{m-2} \text{tr}(\Sigma \Sigma^*(s, Y_s^x) - \Sigma \Sigma^*(s, Y_s^{x'}))\right) ds
+ e^{\lambda s}\left(m | Y_s^x |^{m-2} \langle Y_s^x - Y_s^{x'}, (\Sigma(s, Y_s^x) - \Sigma(s, Y_s^{x'})) \rangle dB_s\right).$$

And now, by taking expectations of the above and using a stopping time argument just as in Lemma 4.3.6 below, we arrive to

$$d\mathbb{E}\left[e^{\lambda s} | Y_s^x - Y_s^{x'} |^m\right] = e^{\lambda s} \left(\lambda \mathbb{E}| Y_s^x - Y_s^{x'} |^m + m \mathbb{E}| Y_s^x - Y_s^{x'} |^{m-2} \langle Y_s^x - Y_s^{x'}, (\beta(s, Y_s^x) - \beta(s, Y_s^{x'})) \rangle\right.
\left. + \frac{m(m-1)}{2} \mathbb{E}| Y_s^x - Y_s^{x'} |^{m-2} \text{tr}(\Sigma \Sigma^*(s, Y_s^x) - \Sigma \Sigma^*(s, Y_s^{x'}))\right) ds. \tag{4.3.3}$$

Using (4.3.2), we obtain

$$d\mathbb{E}\left[e^{\lambda s} | Y_s^x - Y_s^{x'} |^m\right] \leq e^{\lambda s} \left((\lambda - (M_1 - M_0)m) \mathbb{E}| Y_s^x - Y_s^{x'} |^m + m\bar{M}\right) ds. \tag{4.3.4}$$

Now, for the particular choice of $\lambda = m(M_1 - M_0) > 0$:

$$d\mathbb{E}\left[e^{m(M_1 - M_0)s} | Y_s^x - Y_s^{x'} |^m\right] \leq m\bar{M} e^{m(M_1 - M_0)s} ds,$$

which by integrating in between 0 and $s$ becomes

$$\mathbb{E}\left[| Y_s^x - Y_s^{x'} |^m\right] \leq e^{-m(M_1 - M_0)s} \left(|x - x'|^m - \frac{\bar{M}}{M_1 - M_0}\right) + \frac{\bar{M}}{M_1 - M_0}. \tag{4.3.5}$$

Before attacking the moments bounds, we extract a relevant property from Assumption 4.3.2.

**Lemma 4.3.5** (Coercivity consequence of the monotonicity assumption). **Under Assumption 4.3.2, there exists** $M_2 : [0, \infty) \rightarrow \mathbb{R}$ **such that for all** $x \in \mathbb{R}^d$, $s \geq 0$:

$$\langle x, \beta(s, x) \rangle + \frac{p^* - 2}{2} |\Sigma(s, x)|^2 \leq M_2(s) - \frac{(M_1 - M_0)}{2} |x|^2. \tag{4.3.5}$$

**Proof.** When taking $x \in \mathbb{R}^d$ and $x_0 = 0$, the Monotonicity Assumption 4.3.2 implies by the use of the Triangle and Young’s (general) inequalities that for all $s \geq 0$,

$$\langle x, \beta(s, x) \rangle + \frac{p^* - 1}{2} |\Sigma(s, x)|^2 - (p^* - 1)|\Sigma(s, x)|\Sigma^*(s, 0) + \frac{p^* - 1}{2} |\Sigma(s, 0)|^2 
\leq \langle x, \beta(s, 0) \rangle - (M_1 - M_0)|x|^2 + \bar{M} 
\leq \bar{M} + \frac{1}{2(M_1 - M_0)} |\beta(s, 0)|^2 - \frac{(M_1 - M_0)}{2} |x|^2.$$
Now, we can proceed with $\Sigma$ just as we did with $\beta$: by Young’s inequality with an arbitrary $\epsilon > 0$,

$$
\langle x, \beta(s, x) \rangle + \frac{p^* - 1}{2}|\Sigma(s, x)|^2 \\
\leq \bar{M} + \frac{1}{2(M_1 - M_0)}|\beta(s, 0)|^2 - \frac{p^* - 1}{2}|\Sigma(s, 0)|^2 + (p^* - 1)\left(\frac{1}{2\epsilon}|\Sigma(s, 0)|^2 + \frac{\epsilon}{2}|\Sigma(s, x)|^2\right) \\
- \frac{M_1 - M_0}{2}|x|^2.
$$

Hence, for the particular choice of $\epsilon = \frac{1}{p^* - 1} > 0$, we define

$$
M_2(s) = \bar{M} + \frac{1}{2(M_1 - M_0)}|\beta(s, 0)|^2 + \frac{(p^* - 1)(p^* - 2)}{2}|\Sigma(s, 0)|^2
$$

for which the following is satisfied:

$$
\langle x, \beta(s, x) \rangle + \frac{p^* - 2}{2}|\Sigma(s, x)|^2 \leq M_2(s) - \frac{(M_1 - M_0)}{2}|x|^2.
$$

\[\square\]

**Lemma 4.3.6.** Under Assumptions 4.3.1 and 4.3.2, the process $(Y^x_s)_{s \geq 0}$ defined by (4.3.1) satisfies the following bounds for any fixed $m \in \mathbb{N}$ such that $2 \leq m \leq (p^* - 1)/2$ and some constant $C > 0$:

$$
\mathbb{E}[|Y^x_s|^m] \leq |x|^m e^{-m\frac{(M_1 - M_0)}{2}s} + \int_0^s M_2(r) e^{-m\frac{(M_1 - M_0)}{2}(r-s)} dr.
$$

Notice that if $M_2(\cdot)$ is bounded uniformly on $[0, \infty)$, then there exists $C > 0$ such that

$$
\mathbb{E}[|Y^x_s|^m] \leq C(1 + |x|^m e^{-m\frac{(M_1 - M_0)}{2}s}).
$$

**Proof Lemma 4.3.6.** Let $\lambda$ be an arbitrary positive constant and $x$ a starting value in $\mathbb{R}^d$. When Itô’s formula is applied to the stochastic process $(e^{\lambda s}|Y^x_s|^m)_{s \geq 0}$, we obtain, following the notation used in Appendix C.3,

$$
d(e^{\lambda s}|Y^x_s|^m) = e^{\lambda s}\left(\lambda |Y^x_s|^m + m|Y^x_s|^{m-2} \langle Y^x_s, \beta(s, Y^x_s) \rangle + \frac{m(m-1)}{2} |Y^x_s|^{m-2} \text{tr}(\Sigma^*(s, Y^x_s))\right) ds \\
+ e^{\lambda s}\left(m|Y^x_s|^{m-2} \langle Y^x_s, ((\Sigma^j(s, Y^x_s), dB_s))_j \rangle \right).
$$

(4.3.6)

Step 1: And now assume for a moment, and we will prove it later, that

$$
\mathcal{Y}_s := \int_0^s |Y^x_t|^{m-2} \langle Y^x_t, ((\Sigma^j(s, Y^x_s), dB_s))_j \rangle dt
$$

is a true martingale. Then, when taking expectations in (4.3.6),

$$
d\mathbb{E}[e^{\lambda s}|Y^x_s|^m] \\
= e^{\lambda s}\left(\lambda \mathbb{E}[|Y^x_s|^m] + m\mathbb{E}[|Y^x_s|^{m-2} \langle Y^x_s, \beta(s, Y^x_s) \rangle] + \frac{m(m-1)}{2}\mathbb{E}[|Y^x_s|^{m-2} \text{tr}(\Sigma^*(s, Y^x_s))]\right) ds,
$$

(4.3.7)

which can be treated with the coercivity condition (4.3.5) in order to obtain

$$
d\mathbb{E}[e^{\lambda s}|Y^x_s|^m] \leq e^{\lambda s}\left(\lambda - m\frac{(M_1 - M_0)}{2}\right) \mathbb{E}[|Y^x_s|^m] + mM_2(s) ds.
$$

(4.3.8)
Which reads for the particular choice of \( \lambda = \frac{m(M_1 - M_0)}{2} > 0 \) as
\[
d\mathbb{E}[e^{m(M_1 - M_0)/2} | Y^X_s] \leq mM_2(s) e^{m(M_1 - M_0)/2} ds.
\]
And when integrating in between 0 and \( s \),
\[
\mathbb{E}[|Y^X_s|] \leq |x|m e^{-m(M_1 - M_0)/2} + \int_0^s mM_2(r) e^{-m(M_1 - M_0)/(s-r)} dr.
\]

Step 2: In this step we check that the integral process \( \mathcal{Y}_s = \int_0^t |Y^X_r|^{-2} (Y^X_r, (\Sigma^j(s, Y^X_r), dB_s))_j \) is a martingale.

Let \( T_n := \inf\{t \geq 0 | |Y^X_t| > n\} \to \infty, n \to \infty \) and apply Itô’s isometry to the stopped process to obtain:
\[
\mathbb{E}\left[ \left( \int_0^{t \wedge T_n} |Y^X_r|^{-2} (Y^X_r, (\Sigma^j(r, Y^X_r), dB_r))_j \right)^2 \right]
\]
\[
= \mathbb{E}\left[ \left( \int_0^t |Y^X_{s \wedge T_n}|^{-2} (Y^X_{s \wedge T_n}, (\Sigma^j(r \wedge T_n, Y^X_{r \wedge T_n}), dB_{r \wedge T_n}))_j \right)^2 \right]
\]
\[
= \mathbb{E}\left[ \int_0^t |Y^X_{r \wedge T_n}|^{2m-2} \|\Sigma^* (r \wedge T_n, Y^X_{r \wedge T_n})\|^2_{HS} dr \right].
\]

(4.3.9)

Note next that the stopped process \( \mathcal{Y}_{t \wedge T_n} \) is a martingale, which implies that the first part of the proof can be applied by substituting the process \( Y_t \) by its stopped version \( Y_{t \wedge T_n} \). Together with \( \Sigma \) having linear growth (since \( \sigma \) has linear growth and by Lemma C.2.5 (which is not affected by our generalization to Assumption 4.3.2) we have \( \sup_{s \geq 0} \mathbb{E}[|X^X_s|^2] \leq C(1 + |x|^2), \forall x \), they imply that for all \( s \in [0, t] \),
\[
\mathbb{E}\left[ |Y^X_{s \wedge T_n}|^{2m-2} \|\Sigma^* (s \wedge T_n, Y^X_{s \wedge T_n})\|^2_{HS} \right]
\]
\[
\leq \mathbb{E}\left[ |x|^{2m} e^{-2m(M_1 - M_0)/2(s \wedge T_n)} + \int_0^{s \wedge T_n} M_2(r) e^{-2m(M_1 - M_0)/(s \wedge T_n - r)} dr \right]
\]
\[
\leq C(1 + |x|^{2m}).
\]

Given this uniform in time, and so in \( n \), boundedness of \( \mathbb{E}[|Y^X_{s \wedge T_n}|^{2m-2} \|\Sigma^* (s \wedge T_n, Y^X_{s \wedge T_n})\|^2_{HS}] \), from (4.3.9) and Fatou’s Lemma, we conclude that
\[
\mathbb{E}\left[ \left( \int_0^{t \wedge T_n} |Y^X_r|^{-2} (Y^X_r, (\Sigma^j(r, Y^X_r), dB_r))_j \right)^2 \right]
\]
\[
\leq \liminf_{n \to \infty} \mathbb{E}\left[ \left( \int_0^{t \wedge T_n} |Y^X_r|^{-2} (Y^X_r, (\Sigma^j(r, Y^X_r), dB_r))_j \right)^2 \right]
\]
\[
\leq Ct(1 + |x|^{2m}) < \infty.
\]

In other words, \( Y_t \) is a martingale and as such \( E[\mathcal{Y}_t] = 0, \forall t \geq 0 \). Hence, the computations is Step 1 are valid for \( Y_t \) for \( t \geq 0 \) and not only \( Y_{t \wedge T_n} \).

Let us now introduce another family of processes, the so called derivative processes (4.3.1). For a detailed glossary and a recap of the matrix notation see Appendix C.3. The first derivative process given for any \( t \geq 0 \) by \( Z_t = Z_t(x) = \partial_x Y^X_t = \partial_x Y^X_t \in \mathbb{R}^{d \times d} \) is well defined.
(see [5, Theorem 4.10]) and satisfies for all \( i, j \in \{1, \ldots, d\} \):

\[
d Z_{ij} = \sum_{k=1}^{d} Z_{ik}^{t} \partial_{x_{k}} \beta_{j}(t, Y_{t}) \, dt + \sum_{k, l=1}^{d} Z_{ik}^{t} \partial_{x_{k}} \Sigma^{jl}(t, Y_{t}) \, dB_{t}^{l}, \quad Z_{0} = I \in \mathbb{R}^{d \times d}. \tag{4.3.10}
\]

The second derivative process, \( \ddot{Z} = \ddot{Z}(x) = \partial_{x} Z(x) = \partial_{x}^{2} Y^{x} \in \mathbb{R}^{d \times d \times d} \), satisfies \( \ddot{Z}_{0} = 0 \in \mathbb{R}^{d \times d \times d} \) and for any \( t \geq 0, \forall i, j \in \{1, \ldots, d\} \),

\[
d \ddot{Z}^{ri}_{t} = \sum_{k=1}^{d} \ddot{Z}_{ik}^{t} \partial_{x_{k}} \beta_{j}(t, Y_{t}) \, dt + \sum_{k, l=1}^{d} \ddot{Z}_{ik}^{t} \ddot{Z}_{kl}^{t} \partial_{x_{k}} \partial_{x_{l}} \beta_{j}(t, Y_{t}) \, dt
\]

\[
+ \sum_{k, l=1}^{d} \ddot{Z}_{ik}^{t} \partial_{x_{k}} \Sigma^{jl}(t, Y_{t}) \, dB_{t}^{l} + \sum_{k_{1}, k_{2}, l=1}^{d} \ddot{Z}_{ik_{1}}^{t} \ddot{Z}_{kl}^{t} \partial_{x_{k_{1}}} \partial_{x_{l}} \Sigma^{jl}(t, Y_{t}) \, dB_{t}^{l}. \tag{4.3.11}
\]

Next we proceed bounding their moments since they lie at the base of the bounds for the derivatives of \( V \) (see Section 4.3.2).

**Lemma 4.3.7.** For any fixed \( t \geq 0 \), let \( \beta(t, \cdot) \) and \( \Sigma(t, \cdot) \) belong to \( C^{1}(\mathbb{R}^{d}) \) and satisfy Assumption 4.3.2. Then, for \( x \in \mathbb{R}^{d} \), \( Y_{s}^{x} \) satisfying (4.3.1) and any \( (t, z) \in [0, \infty) \times \mathbb{R}^{d} \times \mathbb{R}^{d} \), the following inequality holds pathwise:

\[
\langle (z, \partial_{x} \beta_{j}(t, Y_{s}^{x})), j \rangle + \frac{p^{*}-1}{2} \langle (z, \partial_{x} \Sigma^{jl}(t, Y_{s}^{x})), jl \rangle \leq -(M_{1} - M_{0})|z|^{2}.
\]

**Proof.** By [91, Proposition 2.20], we know that whenever Assumption 4.3.2 holds, for all \( s \geq 0 \) and \( \bar{y} \in \mathbb{R}^{d} \) we have the following

\[
\sup_{z \neq 0} \frac{\langle z, (\partial_{x} \beta_{j}(s, \bar{y})) \rangle + \frac{p^{*}-1}{2} \langle (z, \partial_{x} \Sigma^{jl}(s, \bar{y})), jl \rangle}{|z|^{2}} \leq \sup_{x \neq y \in \mathbb{R}^{d}} \frac{\langle x - y, \beta(s, x) - \beta(s, y) \rangle + \frac{p^{*}-1}{2} \langle \Sigma(s, x) - \Sigma(s, y) \rangle}{|x - y|^{2}}.
\]

Hence, the following inequality holds pathwise

\[
\sup_{z \neq 0} \frac{\langle z, (\partial_{x} \beta_{j}(s, Y_{s}^{x})) \rangle + \frac{p^{*}-1}{2} \langle (z, \partial_{x} \Sigma^{jl}(s, Y_{s}^{x})), jl \rangle}{|z|^{2}} \leq \sup_{x \neq y} \frac{\langle x - y, \beta(s, x) - \beta(s, y) \rangle + \frac{p^{*}-1}{2} \langle \Sigma(s, x) - \Sigma(s, y) \rangle}{|x - y|^{2}}
\]

\[
\leq \sup_{x \neq y} \frac{-(M_{1} - M_{0})|x - y|^{2} + \bar{M}}{|x - y|^{2}} = -(M_{1} - M_{0}) < 0. \quad (4.3.12)
\]

\[\square\]

**Lemma 4.3.8.** Let Assumptions 4.3.1 and 4.3.2 hold. Then, the process \( (Z_{s})_{s \geq 0} \) defined by (4.3.10) satisfies for any \( m \in \mathbb{N} \) such that \( 4 \leq 2m \leq (p^{*}-1)/2 \) the following bound in \( L^{2,2m}_{2,2m} \)-norm:

\[
\mathbb{E}[||Z_{s}||_{2,2m}^{m}] \leq C d e^{-m(M_{1} - M_{0})s}, \quad \forall s \geq 0.
\]
Proof. Let $\lambda$ be an arbitrary positive constant. When Itô’s formula is applied to the process $(e^{\lambda s}|Z^i_s|^m)_{s \geq 0}$, we get:

$$
\begin{align*}
    d(e^{\lambda s}|Z^i_s|^m) &= e^{\lambda s}\left(|Z^i_s|^m + m|Z^i_s|^{m-2}\sum_{j,k} Z^{ij}_s Z^{ik}_s \partial_x \beta_j(s, Y^x_s) \right) \\
    &+ \frac{m(m-1)}{2}|Z^i_s|^{m-2}\sum_{k,l} Z^{ik}_s \partial_x \Sigma^{jl}_s(s, Y^x_s) dB^l_s \\
    &+ m e^{\lambda s}|Z^i_s|^{m-2}\sum_{k,l} Z^{ij}_s Z^{ik}_s \partial_x \Sigma^{jl}_s(dB^l_s) \\
&= e^{\lambda s}\left(|Z^i_s|^m + m|Z^i_s|^{m-2}\sum_{j,l} \langle Z^i_s, \partial_x \Sigma^{jl}_s(s, Y^x_s) \rangle dB^l_s \right) \\
    &+ \frac{m(m-1)}{2}|Z^i_s|^{m-2}\sum_{j,l} \langle Z^i_s, \partial_x \Sigma^{jl}_s(s, Y^x_s) \rangle dB^l_s, \\
\end{align*}
$$

And now, by taking expectations of the above and after a stopping time argument such as the one in the previous proof of Lemma 4.3.6, we arrive to

$$
\begin{align*}
    d\mathbb{E}[e^{\lambda s}|Z^i_s|^m] &= e^{\lambda s}\left(\mathbb{E}[|Z^i_s|^m] + m\mathbb{E}[|Z^i_s|^{m-2}\sum_{j,l} \langle Z^i_s, \partial_x \Sigma^{jl}_s(s, Y^x_s) \rangle \rangle dB^l_s \right) \\
    &+ \frac{m(m-1)}{2}\mathbb{E}[|Z^i_s|^{m-2}\sum_{j,l} \langle Z^i_s, \partial_x \Sigma^{jl}_s(s, Y^x_s) \rangle dB^l_s] ds \\
\end{align*}
$$

Next by Lemma 4.3.7:

$$
\begin{align*}
    d\mathbb{E}[e^{\lambda s}|Z^i_s|^m] &\leq e^{\lambda s}\left(\mathbb{E}[|Z^i_s|^m] - m(M_1 - M_0)\mathbb{E}[|Z^i_s|^m] \right) ds \\
    &= e^{\lambda s}(\lambda - m(M_1 - M_0))\mathbb{E}[|Z^i_s|^m], \\
\end{align*}
$$

which for the particular choice of $\lambda = m(M_1 - M_0) > 0$ implies that

$$
    d\mathbb{E}[e^{m(M_1 - M_0)s}|Z^i_s|^m] \leq 0 \, ds.
$$

Moreover, by integrating in between 0 and $s$ we arrive to

$$
    \mathbb{E}[|Z^i_s|^m] \leq e^{-m(M_1 - M_0)s}.
$$

Finally, when summing over $i$ we obtain that the expectation of the $L_{2,m}$ norm of the random matrix $Z_s$ is bounded uniformly in $s$:

$$
    \mathbb{E}\left[\|Z_s\|_{2,m}^m \right] \leq Cde^{-m(M_1 - M_0)s}, \quad \forall s \geq 0,
$$

for some $C > 0$ which does not depend on $s$ or $x$.  

\square

Lemma 4.3.9. Assume that Assumption 4.3.1 and 4.3.2 hold. Then, $(\tilde{Z}_s)_{s \geq 0}$ defined by (4.3.11) satisfies, for any $m \in \mathbb{N}$ such that $8 \leq 4m^2 \leq (p^* - 1)/2$, the following bound in $L_{2,2m,2m}$-norm
for some constant $C > 0$:

$$E[|\tilde{Z}_s|^m] \leq C d^2 e^{-\frac{(M_1 - M_0)}{s}}, \quad \forall s \geq 0.$$

Proof. Let $\lambda$ be an arbitrary positive constant. When Itô’s formula is applied to the process $(e^{\lambda s}|\tilde{Z}_s|^m)_{s \geq 0}$:

$$d(e^{\lambda s}|\tilde{Z}_s|^m) = e^{\lambda s} \left( |\tilde{Z}_s|^m + m|\tilde{Z}_s|^m \sum_{j,k} \tilde{Z}_{s}^{r_{ij}} \tilde{Z}_{s}^{r_{jk}} \partial_{x_k} \beta_j(s,Y_s) \right)$$

$$+ \frac{m(m-1)}{2} |\tilde{Z}_s|^m \sum_{k_1, k_2, l} \tilde{Z}_{s}^{r_{ik_1}} \tilde{Z}_{s}^{r_{i k_2}} \partial_{x_{k_1}} \Sigma^{j l}(s,Y_s) \tilde{Z}_{s}^{r_{i k_2}} \partial_{x_{k_2}} \Sigma^{j l}(s,Y_s)$$

$$+ e^{\lambda s} \left( m|\tilde{Z}_s|^m \sum_{k_1, k_2, l} \tilde{Z}_{s}^{r_{i k_1}} \tilde{Z}_{s}^{r_{i k_2}} \partial_{x_{k_1}} \Sigma^{j l}(s,Y_s) \tilde{Z}_{s}^{r_{i k_2}} \partial_{x_{k_2}} \Sigma^{j l}(s,Y_s) \right)$$

And now, by taking expectations of the above and using a similar stopping time argument as in Lemma 4.3.6, we arrive to

$$dE[e^{\lambda s}|\tilde{Z}_s|^m] = e^{\lambda s} \left( \lambda |\tilde{Z}_s|^m + mE \left[ |\tilde{Z}_s|^m \sum_{j,k} \tilde{Z}_{s}^{r_{ij}} \tilde{Z}_{s}^{r_{jk}} \partial_{x_k} \beta_j(s,Y_s) \right] \right)$$

$$+ \frac{m(m-1)}{2} E \left[ |\tilde{Z}_s|^m \sum_{k_1, k_2, l} \tilde{Z}_{s}^{r_{ik_1}} \tilde{Z}_{s}^{r_{i k_2}} \partial_{x_{k_1}} \Sigma^{j l}(s,Y_s) \tilde{Z}_{s}^{r_{i k_2}} \partial_{x_{k_2}} \Sigma^{j l}(s,Y_s) \right]$$

$$+ e^{\lambda s} \left( mE \left[ |\tilde{Z}_s|^m \sum_{k_1, k_2, l} \tilde{Z}_{s}^{r_{ik_1}} \tilde{Z}_{s}^{r_{i k_2}} \partial_{x_{k_1}} \Sigma^{j l}(s,Y_s) \tilde{Z}_{s}^{r_{i k_2}} \partial_{x_{k_2}} \Sigma^{j l}(s,Y_s) \right] \right).$$

And so, using the coercivity condition proved in Lemma 4.3.5, we arrive to

$$dE[e^{\lambda s}|\tilde{Z}_s|^m] \leq e^{\lambda s} \left( \lambda - (M_1 - M_0) m \right) E \left[ |\tilde{Z}_s|^m \right]$$

$$+ \frac{m(m-1)}{2} E \left[ |\tilde{Z}_s|^m \sum_{k_1, k_2, l} \tilde{Z}_{s}^{r_{ik_1}} \tilde{Z}_{s}^{r_{i k_2}} \partial_{x_{k_1}} \Sigma^{j l}(s,Y_s) \tilde{Z}_{s}^{r_{i k_2}} \partial_{x_{k_2}} \Sigma^{j l}(s,Y_s) \right].$$

(4.3.13)

Now, on one hand by applying Young’s inequality for an arbitrary $\epsilon > 0$: $ab \leq \frac{p}{p} a^p + \frac{1}{qq} b^q$,
with 
$p = m/(m - 2), q = m/2, a = |Z_s^r|^m - 2$ and

\[ b = \left| \langle \tilde{Z}_s^r, (Z_s^r, (\partial x, (\partial x_{k_1} \beta_j)))(k_1) \rangle + \frac{(m - 1)}{2} \left\| \langle (Z_s^r, ((Z_s^i, \partial x_{x_k} \Sigma^j_l))_k_2) \rangle_{j,l} \right\|_{HS}^2 \right| \frac{(m - 1)}{2} \left\| \langle (Z_s^r, ((Z_s^i, (\partial x_{x_k} \Sigma^j_l))_k_2) \rangle_{j,l} \right\|_{HS}^2 \left\| \langle (Z_s^r, (Z_s^i, (\partial x_{x_k} \Sigma^j_l))_k_2) \rangle_{j,l} \right\|_{HS}^2 \] 

we conclude:

\[ E \left[ |\tilde{Z}_s^r|^m - 2 \langle (\tilde{Z}_s^r, (Z_s^r, (\partial x, (\partial x_{k_1} \beta_j)))(k_1) \rangle \right] \leq \frac{m}{m - 2}\frac{m}{m} \left( \frac{(m - 2)}{m} \right) \left( \left\| \langle (Z_s^r, ((Z_s^i, (\partial x_{x_k} \beta_j)))(k_1) \rangle \right\|_{HS}^2 \right) \leq \frac{m}{m - 2}\frac{m}{m} \left( \left\| \langle (Z_s^r, ((Z_s^i, (\partial x_{x_k} \beta_j)))(k_1) \rangle \right\|_{HS}^2 \right) \]

Let us name $A_1 := \left( \frac{(m - 2)}{m} \right) \left( \left\| \langle (Z_s^r, ((Z_s^i, (\partial x_{x_k} \beta_j)))(k_1) \rangle \right\|_{HS}^2 \right)$. By the algebraic inequality: given any $a, b, c > 0$, there exists $c_m > 0$ such that $(a + b + c)^{m/2} \leq 2^m (a^{m/2} + b^{m/2} + c^{m/2})$,

\[ A_2 := E \left[ \left\| \langle (Z_s^r, ((Z_s^i, (\partial x_{x_k} \beta_j)))(k_1) \rangle \right\|_{HS}^2 \right] \]

Because they are very similar, let us omit the computation of the term

\[ B_1 := \frac{(m - 1)^2}{2} E \left[ \left\| \langle (Z_s^r, ((Z_s^i, (\partial x_{x_k} \beta_j)))(k_1) \rangle \right\|_{HS}^2 \right] \]

and focus on the slightly more complex term

\[ B_2 := 2^m E \left[ \left\| \langle (Z_s^r, ((Z_s^i, (\partial x_{x_k} \beta_j)))(k_1) \rangle \right\|_{HS}^2 \right] \]

112
For the first term of the latter, we can use Young's inequality in the following way: $ab \leq \frac{\varepsilon}{2}a^2 + \frac{1}{2\varepsilon}b^2$ with $a = |Z'_i|^2, b = |\langle Z'_i, i(Z'_i, \partial_x (\partial_x \beta) k) \rangle|^{m/2}$ and $\varepsilon'$ such that 

\[
\frac{m-2}{m} e^{\varepsilon'/m} + \frac{2}{m} e^{\varepsilon'/m - 2} = m(M_1 - M_0) \quad \text{(choice which will be justified later).}
\]

This together with Cauchy–Schwarz applied to the second term of $B_2$ imply

\[
B_2 \leq \frac{2^{m+1}e'}{2} \mathbb{E}[|Z'_i|^m] + \frac{2^{m+1}e'}{2} \mathbb{E}[\langle Z'_i, \langle Z'_i, \partial_x (\partial_x \beta) k \rangle \rangle]^{m/2} 
\]

\[
= \frac{2^{m+1}e'}{2} \mathbb{E}[|Z'_i|^m] + \frac{2^{m+1}e'}{2} \mathbb{E}[\langle Z'_i, \partial_x (\partial_x \beta) k \rangle]^{m/2} 
\]

\[
= \frac{m-2}{m} e^{\varepsilon'/m} |Z'_i|^m + \frac{2}{m} e^{\varepsilon'/m - 2} |Z'_i|^m 
\]

After another Cauchy-Schwarz inequality application, and because of the previous bounds for the moments of $Y, Z$ (see Lemmas 4.3.6 and 4.3.8) up until again order $4m$, together with polynomial growth of the second derivatives of the coefficients, we can continue:

\[
B_2 \leq \frac{2^{m+1}e'}{2} \mathbb{E}[|Z'_i|^m] + C e^{-2m(M_1 - M_0)s} 
\]

Although we omit the details here, by the same trick we obtain

\[
B_1 \leq \frac{2^{m+1}e'}{2} \mathbb{E}[|Z'_i|^m] + C e^{-2m(M_1 - M_0)s} 
\]

Therefore,

\[
A_2 \leq B_1 + B_2 \leq 2^{m+1}e' \mathbb{E}[|Z'_i|^m] + C e^{-2m(M_1 - M_0)s} \, ds 
\]

All together,

\[
A_1 + \frac{2}{me^{m/2}} A_2 \leq \left( \frac{m-2}{m} e^{\varepsilon'/m} |Z'_i|^m + \frac{2}{me^{m/2}} \frac{2^{m+1}e'}{2} \mathbb{E}[|Z'_i|^m] + C e^{-2m(M_1 - M_0)s} \right) ds 
\]

Recall now the specific choice of $\varepsilon'$, this leads to:

\[
A_1 + \frac{2}{me^{m/2}} A_2 \leq m(M_1 - M_0) \mathbb{E}[|Z'_i|^m] + C e^{-2m(M_1 - M_0)s} \, ds 
\]

These imply, when taken back to (4.3.13), that:

\[
d|e^{\lambda s} |Z'_i|^m | \leq e^{\lambda s} \left( C e^{-2m(M_1 - M_0)s} + \left( \frac{m(M_1 - M_0)}{2} \right) \mathbb{E}[|Z'_i|^m] \right) ds 
\]

Now, for the particular choice of $\lambda = m(M_1 - M_0)/2$:

\[
d|e^{\frac{m(M_1 - M_0)}{2} s} |Z'_i|^m | \leq C e^{-3/2m(M_1 - M_0)s} \, ds,
\]
which by integrating in between 0 and $s$ becomes

$$E[|Z_s^r|^m] \leq Ce^{\frac{m(M_1-M_0)}{2}s}.$$

And so, when summing over $r, i$ we have the following uniform bound for some $C$:

$$E[||Z_s||_{2,m,m}^m] \leq Cd^2 e^{\frac{m(M_1-M_0)}{2}s}.$$

\[\Box\]

### 4.3.2 Space derivatives of the backward Kolmogorov function

Now we are ready to bound the derivatives of the solution to the backward Kolmogorov equation (backward Kolmogorov function) introduced in Section 4.1.1.

**Proposition 4.3.10.** Under Assumptions 4.1.2, 4.3.1 and 4.3.2 (holding with $p^* \geq \max \{65, 4l_t + 1, 32l_t + 1\}$), consider $V(t, x) = E[\phi(Y_t^x)]$, where $(Y_t^x)_{t \geq 0}$ is the solution to (4.3.1). Then, there exists $C > 0$ such that for all $(t, x) \in [0, \infty) \times \mathbb{R}^d$ and $n = 1, 2, 3, 4$:

$$||\partial_x^n V(t, x)|| \leq Cd^{n/2} e^{\frac{M(M_1-M_0)+2n}{4}t} \left(1 + \max_{0 \leq i \leq n} \left|\frac{|l|}{l_i}\right| \exp(-\min_{0 \leq i \leq n} \{l_i\}(M_1-M_0)t) \right) + \left(\int_0^t M_1(r) \exp(-\min_{0 \leq i \leq n} \{l_i\}(M_1-M_0)(t-r))dr \right)^{1/2},$$

where the norm represents $|| \cdot ||_2, || \cdot ||_{HS}, || \cdot ||_{2,2,2}$ and $|| \cdot ||_{2,2,2,2}$ as appropriate.

**Proof.** Because of Leibniz formula (see [18, Theorem 12.14] and recall that $\phi$ is continuous with three continuous derivatives), the Chain Rule, the polynomial growth of the first derivative of $\phi$ and the bounds obtained for the processes $Y$ and $Z$ (see Lemmas 4.3.6 and 4.3.8), we know that there exists $C > 0$ such that for any $x \in \mathbb{R}^d$, $e_i$ an element in the canonical basis $\{e_i\}_{i \in \{1, ..., d\}}$ and $h > 0$:

$$\limsup_{h \to 0} \frac{V(t, x) - V(t, x + he_i)}{h} = E\left[\lim_{h \to 0} \frac{\phi(Y_t^x) - \phi(Y_t^{x+he_i})}{h}\right] = E[\langle \partial_x \phi(Y_t^x), (Z_t^{ij})_j \rangle]$$

$$\leq (E[||\partial_x \phi(Y_t^x)||^2])^{1/2} (E[||Z_t||^2_{HS}])^{1/2}$$

$$\leq C(E[1 + |Y_t^x|^{2l_t}])^{1/2} (E[||Z_t||^2_{HS}])^{1/2} < \infty.$$

Since the same argument can be repeated with $(-h)$ and the $i \in \{1, ..., d\}$ was arbitrary, we conclude that the first space derivative of $V$ exists and it is

$$\partial_x V(t, x) = E[\langle \partial_x \phi(Y_t^x), (Z_t^{ij})_j \rangle] = E[\partial_x \phi(Y_t^x)Z_t]$$

(the last equality being notation we introduced). Analogously, but this time because of the second derivatives of $\phi$ is assumed to have order $0 \leq 2l_2 \leq (p^* - 1)/2$ and we also obtained bounds for the processes $\hat{Z}$ (see Lemma 4.3.9), we know that for any $x \in \mathbb{R}^d$, $i, j \in \{1, ..., d\}$ and
again \( e_j \) an element in the canonical basis \( \{ e_i \}_{i \in \{1, \ldots, d\}} \) and \( h > 0 \):

\[
\limsup_{h \to 0} \frac{\partial_{x_i} V(t, x) - \partial_{x_i} V(t, x + he_j)}{h} = \mathbb{E} \left[ \lim_{h \to 0} \frac{\langle \partial_{x}\phi(Y^x_t), (Z^{i_k}_t(x))_k \rangle - \langle \partial_{x}\phi(Y^{x+he}_t), (Z^{i_k}_t(x + he_j))_k \rangle}{h} \right]
\]

\[
= \mathbb{E} \left[ \lim_{h \to 0} \frac{\langle \partial_{x}\phi(Y^x_t), (Z^{i_k}_t(x))_k \rangle - \langle \partial_{x}\phi(Y^{x+he}_t), (Z^{i_k}_t(x + he_j))_k \rangle}{h} \right.
\]

\[
+ \left. \langle \partial_{x}\phi(Y^x_t), (Z^{i_k}_t(x + he_j))_k \rangle - \langle \partial_{x}\phi(Y^{x+he}_t), (Z^{i_k}_t(x + he_j))_k \rangle \right] \frac{1}{h} \right]
\]

\[
= \mathbb{E} \langle \partial_{x}\phi(Y^x_t), (\bar{Z}^{i_k}_t)_k \rangle + \langle \partial_{x_j}\partial_{x}\phi(Y^x_t), (Z^{i_k}_t)_k \rangle \leq (\mathbb{E}[1 + |Y^x_t|^2])^{1/2}(\mathbb{E}[||Z_t||_{HS}^2])^{1/2} + (\mathbb{E}[1 + |Y^x_t|^2])^{1/2}(\mathbb{E}[||Z_t||_{HS}^2])^{1/2} < \infty,
\]

This again implies that the limit exists and it is \( \partial^2_{x^2} V(t, x) = \mathbb{E}[\partial_{x}\phi(Y^x_t) \bar{Z}_t + \partial^2_{x^2}\phi(Y^x_t) Z_t] \).

Moreover, Cauchy–Schwarz inequity, polynomial growth of the first and second derivatives of \( \phi \) and now the explicit bounds of the moments of \( Y, Z \) (see Lemmas 4.3.6, 4.3.8), we obtain:

\[
|\partial_{x} V(t, x)| = |\mathbb{E}[\partial_{x}\phi(Y^x_t) Z_t]|
\]

\[
\leq (\mathbb{E}[||\partial_{x}\phi(Y^x_t)||^2])^{1/2}(\mathbb{E}[||Z_t||_{HS}^2])^{1/2} \leq C(1 + \mathbb{E}[||Y^x_t||^2])^{1/2}(\mathbb{E}[||Z_t||_{HS}^2])^{1/2}
\]

\[
\leq C\left(1 + |x|^2 e^{-l_1(M_1-M_0)t} + \int_0^t M_2(r)e^{-l_1(M_1-M_0)(t-r)} dr \right)^{1/2} \left(e^{-\frac{M_1-M_0}{2} t}\right)^{1/2}
\]

\[
\leq C d^{1/2} e^{-\frac{M_1-M_0}{2} t} \left(1 + |x|e^{-l_1(M_1-M_0)t} + \left(\int_0^t M_2(r)e^{-l_1(M_1-M_0)(t-r)} dr \right)^{1/2} \right).
\]

Analogously, from the additional bounds on the moments of \( \bar{Z} \) (see Lemma 4.3.9),

\[
||\partial^2_{x^2} V(t, x)||_{HS}
\]

\[
\leq \mathbb{E}[||\partial^2_{x^2}\phi(Y^x_t) \bar{Z}_t||_{HS}] + \mathbb{E}[||\partial_{x}\phi(Y^x_t) \bar{Z}_t||_{HS}]
\]

\[
\leq C d^{1/2} e^{-\frac{M_1-M_0}{4} t} \left(1 + |x|^2 e^{-l_2(M_1-M_0)t} + \int_0^t M_2(r)e^{-l_2(M_1-M_0)(t-r)} dr \right)^{1/2}
\]

\[
+ \left(\int_0^t M_2(r)e^{-l_1(M_1-M_0)(t-r)} dr \right)^{1/2} \left(e^{-\frac{M_1-M_0}{2} t}\right)^{1/2}
\]

\[
\leq C d e^{-\frac{M_1-M_0}{4} t} \left(1 + |x|^{max{l_1,l_2}} e^{-min{l_1,l_2}(M_1-M_0)t}
\]

\[
+ \left(\int_0^t M_2(r)e^{-min{l_1,l_2}(M_1-M_0)(t-r)} dr \right)^{1/2} \right).
\]

Finally, for the third derivative existence we argue in the same way and moreover, from the additional bounds on the moments of the third derivative process \( \bar{Z} \) (see Lemma C.3.1) we
know that there exist $C, C_1 > 0$ such that (again in our loose notation):

$$||\partial_x^3 V(t, x)||_{2, 2, 2}$$

$$\leq \mathbb{E}[||\partial_x^3 \phi(Y_t^x) Z_t Z_t||_{2, 2, 2}] + \mathbb{E}[||\partial_x^2 \phi(Y_t^x) \tilde{Z}_t||_{2, 2, 2}] + \mathbb{E}[||\partial_x \phi(Y_t^x) \tilde{Z}_t||_{2, 2, 2}] + \mathbb{E}[||\partial_x \phi(Y_t^x) \tilde{Z}_t||_{2, 2, 2}]$$

$$\leq C d^{3/2} e^{-C_1 t} \left(1 + |x|^{\max\{l_1, l_2, l_3\}} e^{-\min\{l_1, l_2, l_3\}(M_1 - M_0) t} + \left(\int_0^t M_2(r) e^{-\min\{l_1, l_2, l_3\}(t-r)} dr\right)^{1/2}\right).$$

The analogous holds under our assumptions for the fourth derivative process. 

**Remark 4.3.11.** Suppose that there exists $\kappa > 0$ such that

$$\sup_{t \geq 0} \int_0^t M_2(r) e^{-\min\{l_1, l_2, l_3\}(M_1 - M_0)(t-r)} dr < \infty$$

and let us denote for the remainder of this chapter unless stated otherwise

$$m^* := 2 \max\{l_1, l_2, l_3, l_4\}.$$ 

Then the bounds above become for some $C, c > 0$ (which may change value but not sign form line to line) the following uniform bounds holding for all $(t, x) \in [0, \infty) \times \mathbb{R}^d$:

$$\max\{||\partial_x V(t, x)||, ||\partial_x^2 V(t, x)||_{HS}, ||\partial_x^3 V(t, x)||_{2, 2, 2}, ||\partial_x^4 V(t, x)||_{2, 2, 2, 2}\} \leq c e^{-\kappa t} (1 + |x|^{m^*/2}).$$

As a final comment, notice that at the expense of a higher regularity assumption on the test function $\phi$ and the monotonicity assumption holding for bigger parameters $p^*$, we could iterate the same idea as with the second derivative process $\tilde{Z}$ even for higher order derivative processes (we include the computation for the third derivative in Lemma C.3.1). This is due to the fact that the leading part in the SDE satisfied by this extra derivative process is still composed from the coefficients $(s, z) \mapsto (z \partial_x \beta_j(y), z \partial_z \Sigma_j(y))$, for which we proved coercivity. The remaining part is going to be based on the polynomial growth of higher order derivatives of the coefficients and moments of higher order derivative processes.

### 4.4 Euler Scheme approximation of the subsampled system

In this section we consider the following SDE:

$$dX_t^\xi = b(X_t^\xi, \mathbb{E}[\alpha(X_t^\xi)], \nu) dt + \sigma(X_t^\xi, \nu) dB_t, \quad \forall t \in [0, \infty); \quad X_0^\xi = \xi. \quad (4.4.1)$$

For this particular structure on the McKean non-linearity, we introduce therefore an additional assumption:

**Assumption 4.4.1.** The drift $b : \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ is linear in the first measure component, i.e. it can be cast into the shape $b : \mathbb{R}^d \times \mathbb{R} \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}^d$, where $y = \mathbb{E}[\alpha(X)], X \sim \mu$ with:

1. $\alpha : \mathbb{R}^d \rightarrow \mathbb{R}$ is a bounded with two bounded derivatives function (i.e $\alpha \in C_b^2(\mathbb{R}^d)$) and
2. The first and second derivatives in the second space component, meaning $\partial_y b(x, y, \nu)$ and $\partial^2_y b(x, y, \nu)$, are bounded uniformly on $\mathbb{R}^d \times \mathbb{R} \times \mathcal{P}(\mathbb{R}^d)$. 

116
Although for now we will be referring to the drift as \( b(x, \mu, \nu) \) and only when the linearity in measure plays a relevant role we underline its significance by identifying \( b(x, \mu, \nu) := b(x, \mathbb{E}[\alpha(X)], \nu), X \sim \mu \), the reader must be aware of the abuse of notation which is made obvious by the change in notation.

Notice that for this section the drift is considerably more general compared to the Section 4.2.1 since our hopes of obtaining a weak error order independent of the subsample are already frustrated in the discrete scenario and there is no need to consider overly simplified drifts. Indeed, recall Remark (4.2.4) where the time discretization already suggests at an intuitive level that the size of the subsample has a more relevant role in the weak error’s order.

Now, in the randomly sub–sampled particle system approximation, for some \( T > 0 \) we have \( \mathbb{E}[\phi(X_T^\xi)] \) approximated, in continuous time and for some \( 0 < S \leq N \in \mathbb{N} \) and \( 0 < S' \leq N' \in \mathbb{N} \), by \( \frac{1}{N} \sum_{j=1}^{N} \phi(Z_{t,j}^{i,N,S}) \), where \( \{Z_{t,j}^{i,N,S}\}_{i=1,...,N} \) satisfy (4.2.1). For convenience, we recall that for \( \xi^i \), \( B_i \) i.i.d copies of \( \xi \) and \( B \) respectively,

\[
\begin{align*}
\left\{ \begin{array}{l}
    dZ_{t}^{i,N,S} = b(Z_{t}^{i,N,S}, \mu_{t}^{i,N,S}, \nu_{t}^{i,N,S}) \, dt + \sigma(Z_{t}^{i,N,S}, \nu_{t}^{i,N,S}) \, dB_{t}^{i}, \quad \forall \, t \in [0, \infty), \, i = 1, ..., N; \\
    Z_{0}^{i,N,S} = \xi_{i},
\end{array} \right.
\end{align*}
\]

where recall that the dependence on \( N', S' \) is not included on the notation for the particle system \( Z \) in order to avoid cumbersome notation.

For the discretization we use an Euler Scheme: fixed a number of steps \( M \in \mathbb{N} \) and step size \( h > 0 \) such that \( T = Mh \), we denote (here and for the rest of this document)

\[
\eta(t) := t_{k-1} = (k-1)h \quad \text{if} \quad t \in [t_{k-1}, t_{k}), \, k = 1, ..., M;
\]

and consider

\[
\begin{align*}
\left\{ \begin{array}{l}
    d\tilde{Z}_{t}^{i,N,S} = b(\tilde{Z}_{\eta(t)}^{i,N,S}, \tilde{\mu}_{\eta(t)}^{i,N,S}, \nu_{\eta(t)}^{i,N,S}) \, dt + \sigma(\tilde{Z}_{\eta(t)}^{i,N,S}, \nu_{\eta(t)}^{i,N,S}) \, dB_{t}^{i}, \quad i = 1, ..., N; \\
    \tilde{Z}_{0}^{i,N,S} = \xi_{i},
\end{array} \right.
\end{align*}
\]

Again, we focus on approximating a linear function of \( \mathcal{L}(X_T) \) for some \( T > 0 \). With that in mind recall, for some \( \phi \) satisfying Assumption 4.1.2, the function \( V(t, x) = \mathbb{E}[\phi(X_T^\xi)] \) and the associated (defined by \( (4.3.1) \)) \( V(t, x; \xi) \in [0, \infty) \times \mathbb{R}^d \). Our main tool is discussed in Section 4.1.1: fixed \( \xi \), \( V(\cdot, \cdot; \xi) \) satisfies the PDE obtained from (4.1.3) after a time reversal: for all \( (t, x) \in [0, \infty) \times \mathbb{R}^d \),

\[
\partial_{x} V(t, x; \xi) - \frac{1}{2} \mathrm{tr}(\sigma(x, \nu) \partial_{x}^{2} V(t, x; \xi) \sigma^{*}(x, \nu)) = \langle \partial_{x} V(t, x; \xi), b(x, \mathcal{L}(X_T^\xi), \nu) \rangle;
\]

\[
V(0, x; \xi) = \phi(x).
\]

Notice that existence and smoothness of solution to this PDE is ensured by [51, Proposition 2.3].

**Theorem 4.4.2.** Fix \( 0 < S \leq N \in \mathbb{N} \) and \( 0 < S' \leq N' \in \mathbb{N} \) and let Assumptions 4.1.1, 4.1.2, 4.1.3, 4.1.4 and 4.1.4 hold. Given \( \xi \), a \( p^{*} \)-integrable random variable, consider \( X_T^\xi \) solution to (4.4.1) and its particle system approximation (4.4.2) with the corresponding Euler discretization (4.4.3). Then, there exists a positive constant \( C \) which is independent of \( h, N, N', S, S' \) or \( T \) such
that
\[
\sup_{T>0} \left| \mathbb{E}[\phi(X_T^c)] - \mathbb{E}\left[ \frac{1}{N} \sum_{j=1}^{N} \phi(\bar{Z}_T^{j,N,S}) \right] \right| \leq C \left( 1/N + h + h^{1/2} \mathcal{R}(S,S') + \mathcal{R}(S,S') \right).
\]

The result above can be thought as a three stages analysis. The first one is the Euler scheme approximation of a McKean-Vlasov SDE and is covered by Lemma C.4.1. The second, presented in Lemma C.5.1, is built on the first result and covers the Euler scheme approximation of particle system coupled through the empirical mean of all their positions. A non uniform version of these bounds have already been proved in [57] and [67] respectively. We include our own proof for uniform estimates in the Appendix. Third and final, we cover the subsampling layer in the proof of Theorem 4.4.2.

Finally, we observe that as a by–product of the proof, we generalize the weak propagation result [50, Theorem B.3] to uniform rate, although we work with linear (in measure) test functions and a more restricted set of coefficients.

**Proof Theorem 4.4.2.** Since the particles are identically distributed, we have
\[
\mathbb{E}\left[ \frac{1}{N} \sum_{j=1}^{N} \phi(\bar{Z}_T^{j,N,S}) \right] = \mathbb{E}[\phi(\bar{Z}_T^{1,N,S})],
\]
and hence the weak error can be rewritten as \( |\mathbb{E}[\phi(X_T^c)] - \mathbb{E}[\phi(Z_T^{1,N,S})]| \).

In this circumstances, the weak error reformulation, Itô’s formula and PDE (4.4.4) imply, similarly to Theorem 2.3, that
\[
|\mathbb{E}[\phi(X_T^c)] - \mathbb{E}[\phi(Z_T^{1,N,S})]| = |\mathbb{E}[V(T, \bar{Z}_T^{1,N,S}) - V(0, \bar{Z}_T^{1,N,S})]| = |\mathbb{E}\left[ \sum_{k=1}^{M} V(t_k, \bar{Z}_T^{1,N,S}) - V(t_k, \bar{Z}_T^{1,N,S}) \right]| = |\mathbb{E}\left[ \sum_{k=1}^{M} \left| \int_{t_{k-1}}^{t_k} \partial_s V(s, \bar{Z}_T^{1,N,S}) - \left\{ \partial_s V(s, \bar{Z}_T^{1,N,S}) \right\} b(\bar{Z}_T^{1,N,S}, \mu_{\eta(s)}, \nu_{\eta(s)}, N') \right| ds \right]|.
\]

And after a series of decompositions:
\[
\leq |\mathbb{E}\left[ \int_{0}^{T} \left\{ \partial_s V(s, \bar{Z}_T^{1,N,S}) \right\} b(\bar{Z}_T^{1,N,S}, \mu_{\eta(s)}, \nu_{\eta(s)}, N') - b(\bar{Z}_T^{1,N,S}, \mu_{\eta(s)}, \nu_{\eta(s)}, N') \right] ds \right]| \quad \text{(4.4.5)}
\]
\[
+ |\mathbb{E}\left[ \int_{0}^{T} \left\{ \partial_s V(s, \bar{Z}_T^{1,N,S}) \right\} b(\bar{Z}_T^{1,N,S}, \mu_{\eta(s)}, \nu_{\eta(s)}, N') - b(\bar{Z}_T^{1,N,S}, \mu_{\eta(s)}, \nu_{\eta(s)}, N') \right] ds \right]| \quad \text{(4.4.6)}
\]

118
Notice first that since the terms (4.4.5) and (4.4.6) are similar in the sense that we are subsampling each of the measure components. With respect to the term (4.4.6), aiming to obtain a bound in $h$, we can rewrite it as:

$$
\begin{align*}
|\mathbb{E}\left[ \int_0^T \langle \partial_x V(s, \tilde{Z}_{1,N,S}^{1,N,S}), b(\tilde{Z}_{1,N,S}^{1,N,S}, \bar{\eta}_{\eta(s)}, \bar{\mu}_{\eta(s)}, \bar{\nu}_{\eta(s)}) - \mathcal{F}(X_s^\xi, v) \rangle ds \right] |
\leq |\mathbb{E}\left[ \int_0^T \langle \partial_x V(s, \tilde{Z}_{1,N,S}^{1,N,S} - \partial_x V(s, \tilde{Z}_{1,N,S}^{1,N,S}), b(\tilde{Z}_{1,N,S}^{1,N,S}, \bar{\eta}_{\eta(s)}, \bar{\mu}_{\eta(s)}, \bar{\nu}_{\eta(s)}) - b(\tilde{Z}_{1,N,S}^{1,N,S}, \bar{\eta}_{\eta(s)}, \bar{\mu}_{\eta(s)}, \bar{\nu}_{\eta(s)}) \rangle ds \right]
+ |\mathbb{E}\left[ \int_0^T \langle \partial_x V(s, \tilde{Z}_{1,N,S}^{1,N,S} - \partial_x V(s, \tilde{Z}_{1,N,S}^{1,N,S}), b(\tilde{Z}_{1,N,S}^{1,N,S}, \bar{\eta}_{\eta(s)}, \bar{\mu}_{\eta(s)}, \bar{\nu}_{\eta(s)}) - b(\tilde{Z}_{1,N,S}^{1,N,S}, \bar{\eta}_{\eta(s)}, \bar{\mu}_{\eta(s)}, \bar{\nu}_{\eta(s)}) \rangle ds \right] |
\end{align*}
$$

On one hand, by the tower property and Assumption 4.1.4 together with the derivatives decay obtained in Proposition 4.3.10 (see also Remark 4.3.11) and the uniform bounds on the moments (see Appendix C.2), there exist $C, c > 0$ such that

$$
\begin{align*}
|\mathbb{E}\left[ \int_0^T \langle \partial_x V(s, \tilde{Z}_{1,N,S}^{1,N,S}), b(\tilde{Z}_{1,N,S}^{1,N,S}, \bar{\eta}_{\eta(s)}, \bar{\mu}_{\eta(s)}, \bar{\nu}_{\eta(s)}) - b(\tilde{Z}_{1,N,S}^{1,N,S}, \bar{\eta}_{\eta(s)}, \bar{\mu}_{\eta(s)}, \bar{\nu}_{\eta(s)}) \rangle ds \right] |
\leq |\mathbb{E}\left[ \int_0^T |\partial_x V(s, \tilde{Z}_{1,N,S}^{1,N,S} - \partial_x V(s, \tilde{Z}_{1,N,S}^{1,N,S}), b(\tilde{Z}_{1,N,S}^{1,N,S}, \bar{\eta}_{\eta(s)}, \bar{\mu}_{\eta(s)}, \bar{\nu}_{\eta(s)}) - b(\tilde{Z}_{1,N,S}^{1,N,S}, \bar{\eta}_{\eta(s)}, \bar{\mu}_{\eta(s)}, \bar{\nu}_{\eta(s)}) \rangle ds \right] |
\leq C\mathcal{R}(S, S')
\end{align*}
$$

Notice here that if the subsampled average is an unbiased estimator of the non-subsampled one (as is the case for drifts with linear dependence on the measure, see (4.2.4)), then this term will cancel altogether. However, in the end it will not affect the total order of the weak error if we do allow a bias of order smaller than $\Theta(\frac{1}{N})$, for e.g $\mathcal{R}(S, S') = \Theta(\frac{1}{S^2} + \frac{1}{S^2})$, where $S, S' \sim \sqrt{\bar{N}}$.

On the other hand, by the Cauchy–Schwarz inequality we obtain:

$$
\begin{align*}
|\mathbb{E}\left[ \int_0^T \langle \partial_x V(s, \tilde{Z}_{1,N,S}^{1,N,S} - \partial_x V(s, \tilde{Z}_{1,N,S}^{1,N,S}), b(\tilde{Z}_{1,N,S}^{1,N,S}, \bar{\eta}_{\eta(s)}, \bar{\mu}_{\eta(s)}, \bar{\nu}_{\eta(s)}) - b(\tilde{Z}_{1,N,S}^{1,N,S}, \bar{\eta}_{\eta(s)}, \bar{\mu}_{\eta(s)}, \bar{\nu}_{\eta(s)}) \rangle ds \right] |
\leq C\mathcal{R}(S, S')
\end{align*}
$$
\[
\leq \int_0^T \left( E \left[ \left| \partial_x V(s, Z_{T-t}^{1,N,S}) - \partial_x V(s, \tilde{Z}_{T-t}^{1,N,S}) \right|^2 \right] \right)^{1/2} \times \left( E \left[ \left| b(\tilde{Z}_{\eta(t)}^{1,N,S}, \mu_{\eta(t)}, V^{\theta_{\eta(t)}}, \eta_i) - b(\tilde{Z}_{\eta(t)}^{1,N,S}, \mu_{\eta(t)}, V) \right|^2 \right] \right)^{1/2} \, ds.
\]

Now notice that when Itô’s formula is applied to \( \partial_x V(t, Z_{T-t}^{1,N,S}) \), \( i \in \{1, \ldots, d\} \), one obtains

\[
\partial_x V(s, Z_{T-t}^{1,N,S}) - \partial_x V(s, \tilde{Z}_{T-t}^{1,N,S}) = \int_{\eta(t)}^s \left( -\partial_t (\partial_x V(t, x), b(x, \mathcal{L}(X_t^c), v)) + \langle \partial_x (\partial_x V(t, x), b(x, \mathcal{L}(X_t^c), v) \rangle + \frac{1}{2} \text{tr}(\sigma(Z_{\eta(t)}^{i,N,S}, \theta_{\eta(t)}^{\theta^e}), \partial_x^2 V(t, x)\sigma^*(x, v) + \frac{1}{2} \text{tr}(\sigma(x, v)\partial_x^2 \partial_x V(t, x)\sigma^*(x, v)); \right.
\]

Moreover, since \( V \in C^{1,3}([0, \infty) \times \mathbb{R}^d) \), one can obtain an analogous PDE to that obtained for \( V \) and which takes the following form

\[
\begin{align*}
\partial_t (\partial_x V)(t, x) &= \langle \partial_x V(t, x), \partial_x b(x, \mathcal{L}(X_t^c), v) \rangle + \langle \partial_x \partial_x V(t, x), b(x, \mathcal{L}(X_t^c), v) \rangle \\
&+ \frac{1}{2} \text{tr}(\sigma(x, v)\partial_x^2 V(t, x)\sigma^*(x, v)); \right.
\]

Notice here the need for the derivative bounds stated in Assumption 4.4.1, since in our case one of the terms in the PDE above becomes:

\[
\langle \partial_x V(t, x), \partial_x b(x, \mathcal{L}(X_t^c), v) \rangle = \langle \partial_x V(t, x), \partial_x \partial_x b(x, \mathcal{L}(X_t^c), v) \rangle \\
= \langle \partial_x V(t, x), \partial_x \partial_x b(x, \mathcal{L}(X_t^c), v) \rangle \right) + \langle \partial_x V(t, x), \partial_x \partial_x b(x, \mathcal{L}(X_t^c), v) \rangle \right) = [\partial_x V(t, x), \partial_x \partial_x b(x, \mathcal{L}(X_t^c), v) \rangle \right).
\]

Using this PDE to simplify equation (4.4.11),

\[
\begin{align*}
\partial_x V(s, Z_{T-t}^{1,N,S}) - \partial_x V(s, \tilde{Z}_{T-t}^{1,N,S}) &= \int_{\eta(t)}^s \langle \partial_x (\partial_x V(t, \tilde{Z}_{T-t}^{1,N,S}), b(\tilde{Z}_{\eta(t)}^{1,N,S}, \frac{1}{S} \sum_{j=1}^S \sigma(Z_{\eta(t)}^{j,N,S}, v^{U_{\eta(t)}}, N^t) - b(\tilde{Z}_{T-t}^{1,N,S}, \mathcal{E}(X_t^c), v) \right) \\
&- \frac{1}{2} \text{tr}(\sigma(Z_{\eta(t)}^{i,N,S}, \theta_{\eta(t)}^{\theta^e}), \partial_x^2 \partial_x V(t, \tilde{Z}_{T-t}^{1,N,S}) \right) \right) + \langle \partial_x V(t, \tilde{Z}_{T-t}^{1,N,S}), \partial_x b(\tilde{Z}_{T-t}^{1,N,S}, \mathcal{E}(X_t^c), v) \rangle \\
&- \langle \partial_x V(t, \tilde{Z}_{T-t}^{1,N,S}), \partial_x b(\tilde{Z}_{T-t}^{1,N,S}, \mathcal{E}(X_t^c), v) \rangle \right) \right) \right) = [\partial_x V(t, x), \partial_x \partial_x b(x, \mathcal{L}(X_t^c), v) \rangle \right).
\]

120
Now by using Itô’s isometry, Hölder’s inequality first and afterwards Assumptions 4.1.1 and 4.4.1 on the Lipschitz character and linear growth of \( \sigma \) and the boundedness of the derivatives of the coefficients, together with the bounds on the first three space derivatives of \( V \) we obtained in Preposition 4.3.10 (see also Remark 4.3.11), there exist \( C, c > 0 \) such that:

\[
\mathbb{E} \left[ \left| \partial_{x_i} V(s, \tilde{Z}_{T-s}^{1,N,S}) - \partial_{x_i} V(\eta(s), \tilde{Z}_{T-\eta(s)}^{1,N,S}) \right|^2 \right] \\
\leq h \mathbb{E} \left[ \int_{\eta(s)}^{s} \left| \partial_x(\partial_{x_i} V)(t, \tilde{Z}_{T-t}^{1,N,S}) \cdot b(\tilde{Z}_{\eta(t)}^{1,N,S}, \frac{1}{S} \sum_{j=1}^{S} \alpha(\tilde{Z}_{\eta(t)}^{\Theta(j),N,S}, \nu)^{U(\eta(t),N')}) - b(\tilde{Z}_{t}^{1,N,S}, \mathbb{E}[\alpha(X_t)], \nu) \right|^2 d\eta(t) \\
\right] + \frac{1}{2} \mathbb{E} \left[ \left| \partial_{x_i} V(t, \tilde{Z}_{T-t}^{1,N,S}) \sigma(\tilde{Z}_{t}^{1,N,S}, \nu^*) + b(\tilde{Z}_{t}^{1,N,S}, \mathbb{E}[\alpha(X_t)], \nu) \right|^2 \right] \\
\left. \cdot \left( \sigma(\tilde{Z}_{\eta(t)}^{1,N,S}, \nu^*) - \sigma(\tilde{Z}_{t}^{1,N,S}, \nu)^* \right) \right] \\
- \mathbb{E} \left[ \int_{\eta(s)}^{s} \left| \partial_{x_i} V(t, \tilde{Z}_{T-t}^{1,N,S}) \sigma(\tilde{Z}_{t}^{1,N,S}, \nu^*) \right|^2 dt \right] \\
+ \mathbb{E} \left[ \int_{\eta(s)}^{s} \left| \partial_{x_i} V(t, \tilde{Z}_{T-t}^{1,N,S}) \sigma(\tilde{Z}_{t}^{1,N,S}, \nu^*) \right|^2 dt \right] \\
\leq h \mathbb{E} \left[ \int_{\eta(s)}^{s} e^{-ct}(1 + |\tilde{Z}_{T-t}^{1,N,S}|^2)^{m^*} \\
\cdot \left| b(\tilde{Z}_{\eta(t)}^{1,N,S}, \frac{1}{S} \sum_{j=1}^{S} \alpha(\tilde{Z}_{\eta(t)}^{\Theta(j),N,S}, \nu)^{U(\eta(t),N')}) - b(\tilde{Z}_{t}^{1,N,S}, \mathbb{E}[\alpha(X_t)], \nu) \right|^2 dt \\
+ e^{-ct}(1 + |\tilde{Z}_{T-t}^{1,N,S}|^{m^*+4}) dt \right] \\
+ CE \left[ \int_{\eta(s)}^{s} e^{-ct}(1 + |\tilde{Z}_{T-t}^{1,N,S}|^2) dt \right].
\]

We continue by using the tower property and compatibility condition again

\[
\leq Ch \int_{\eta(s)}^{s} e^{-ct} \mathbb{E} \left[ (1 + |\tilde{Z}_{T-t}^{1,N,S}|^{m^*}) \mathbb{E}[\tilde{Z}_{\eta(t)}^{1,N,S}] \left| b(\tilde{Z}_{\eta(t)}^{1,N,S}, \frac{1}{S} \sum_{j=1}^{S} \alpha(\tilde{Z}_{\eta(t)}^{\Theta(j),N,S}, \nu)^{U(\eta(t),N')}) - b(\tilde{Z}_{t}^{1,N,S}, \mathbb{E}[\alpha(X_t)], \nu) \right|^2 \right] dt \\
+ e^{-ct} \mathbb{E} \left[ (1 + |\tilde{Z}_{T-t}^{1,N,S}|^{m^*+4}) \right] dt + C \int_{\eta(s)}^{s} e^{-ct} dt.
\]

Next, after an extra decomposition, Assumption 4.1.4 over the variance of the bias together with the growth condition for \( b \) (see Remark 1 above inequality (4.1.8)) and finally uniform in time boundedness of moments (see Appendix C.2) allow us to conclude that:

\[
\leq Ch \int_{\eta(s)}^{s} e^{-ct} \mathbb{E} \left[ (1 + |\tilde{Z}_{T-t}^{1,N,S}|^{m^*}) \mathbb{E}[\tilde{Z}_{\eta(t)}^{1,N,S}] \left| b(\tilde{Z}_{\eta(t)}^{1,N,S}, \frac{1}{S} \sum_{j=1}^{S} \alpha(\tilde{Z}_{\eta(t)}^{\Theta(j),N,S}, \nu)^{U(\eta(t),N')}) - b(\tilde{Z}_{t}^{1,N,S}, \mathbb{E}[\alpha(X_t)], \nu) \right|^2 \right] dt
\]
\[ + e^{-ct} \mathbb{E}\left[ \left( 1 + |\hat{Z}_{T-t}^{i, N, S}|^m \right) \mathbb{E}[\hat{Z}_{\eta(t)}] \left[ b(\hat{Z}_{\eta(t)}^{i, N, S}, \mathbb{E}[\alpha(X)], v) - b(\hat{Z}_{\eta(t)}^{i, N, S}, \mathbb{E}[\alpha(X)], v) \right]^2 \right] \]
\[ + e^{-ct} \mathbb{E}[1 + |\hat{Z}_{T-t}^{i, N, S}|^{m+4}] dt + C \int_{\eta(s)}^s e^{-ct} dt \]
\[ \leq C h \int_{\eta(s)}^s e^{-ct} \left[ 1 + |\hat{Z}_{T-t}^{i, N, S}|^m \right] e^{-ct} dt + C \int_{\eta(s)}^s e^{-ct} dt \]
\[ \leq C(1 + h + \mathcal{R}(S, S')) \] (4.4.13)

Hence, the above combined with the assumed control on the variance of the bias (see Assumption 4.1.4) and again the uniform in time bounds of the moments, imply after applying Cauchy–Schwarz that:
\[ \mathbb{E}\left[ \int_0^T \left( \partial_s V(s, \hat{Z}_{T-s}^{i, N, S}) - \partial_s V(s, \hat{Z}_{T-\eta(s)}^{i, N, S}) \right) b(\hat{Z}_{\eta(s)}^{i, N, S}, \mathbb{E}[\alpha(X)], v) \mathbb{E}[\hat{Z}_{\eta(s)}^{i, N, S}, v] ds \right] \]
\[ \leq \int_0^T \left( \mathbb{E}\left[ \left( \partial_s V(s, \hat{Z}_{T-s}^{i, N, S}) - \partial_s V(s, \hat{Z}_{T-\eta(s)}^{i, N, S}) \right)^2 \right] \right)^{1/2} \left( \mathcal{R}(S, S') \right)^{1/2} ds \]
\[ \leq C \left( \mathcal{R}(S, S') \right)^{1/2} \int_0^T h^{1/2} e^{-c/2 \eta(s)} (1 + h^2 + \mathcal{R}(S, S'))^{1/2} ds \]
\[ \leq C h^{1/2} \left( \mathcal{R}(S, S') \right)^{1/2} (1 + h + \mathcal{R}(S, S'))^{1/2} \]
\[ \leq C h^{1/2} \mathcal{R}(S, S'), \]

where we stress again that \( C \) is a positive constant not depending on the time \( T \) or the batch size. Finally, we can put together (4.4.10) and (4.4.13) to obtain that the term (4.4.6) is bounded by \( C(\mathcal{R}(S, S') + h^{1/2} \mathcal{R}(S, S')) \).

Again, notice that the computation for the term (4.4.5) would be simplified for the term (4.4.6) since we are keeping fixed the space component and internal measure component rather than the external measure component. In particular, since we do not need to arrive to a bound on which to apply Grönwall’s Lemma, the linear dependence in the external measure component is unnecessary. So this time we need to account for the error induced by using a subsampled version of the empirical measure rather than the whole empirical measure of the particle system. The conclusion is once again relying on an analogous version of (4.4.10) and (4.4.13):
\[ \mathbb{E}\left[ \int_0^T \left( \partial_s V(s, \hat{Z}_{T-s}^{i, N, S}) - b(\hat{Z}_{\eta(s)}^{i, N, S}, \mathbb{E}[\alpha(X)], v) - b(\hat{Z}_{\eta(s)}^{i, N, S}, \mathbb{E}[\alpha(X)], v) \right) \bigg| \mathcal{F}_s \right] ds \]
\[
\leq C\mathcal{R}(S,S')+ C\int_0^T \left( \mathbb{E}[|\partial_x V(s, Z_{T-s}^{1,N,S}) - \partial_x V(s, \tilde{Z}_{T-s}^{1,N,S})|^2] \right)^{1/2} \\
\times \left( \mathbb{E}[|b(\tilde{Z}_{\eta(s)}^{1,N,S}, \tilde{Z}_{\eta(s)}^{N,\Theta}, \Theta_{\eta(s)}^{\Theta}) - b(\tilde{Z}_{\eta(s)}^{1,N}, \tilde{Z}_{\eta(s)}^{N,S}, \tilde{v}_{\eta(s)}^{\Theta})|^2] \right)^{1/2} \, ds \\
\leq C\mathcal{R}(S,S') + h^{1/2} \mathcal{R}(S,S').
\]

Next, the term \((4.4.8)\) can be decomposed as follows:

\[
\mathbb{E} \left[ \int_0^T \frac{1}{2} \text{tr} \left( \sigma(\tilde{Z}_{\eta(s)}^{1,N,S}, V_{\eta(s)}^{\Theta}) \partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S}) \sigma^*(\tilde{Z}_{\eta(s)}^{1,N,S}, V_{\eta(s)}^{\Theta}) \\
- \sigma(\tilde{Z}_{\eta(s)}^{1,N,S}, V)(\partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S}) - \partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S})) \sigma^*(\tilde{Z}_{\eta(s)}^{1,N,S}, V_{\eta(s)}^{\Theta}) \right) \, ds \right] \\
\leq \mathbb{E} \left[ \int_0^T \frac{1}{2} \text{tr} \left( \sigma(\tilde{Z}_{\eta(s)}^{1,N,S}, V_{\eta(s)}^{\Theta}) (\partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S}) - \partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S})) \sigma^*(\tilde{Z}_{\eta(s)}^{1,N,S}, V_{\eta(s)}^{\Theta}) \right) \, ds \right] \\
+ \mathbb{E} \left[ \int_0^T \frac{1}{2} \text{tr} \left( \sigma(\tilde{Z}_{\eta(s)}^{1,N,S}, V_{\eta(s)}^{\Theta}) - \sigma(\tilde{Z}_{\eta(s)}^{1,N,S}, V) \right) \partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S}) \right. \\
\left. \times \left( \sigma(\tilde{Z}_{\eta(s)}^{1,N,S}, V_{\eta(s)}^{\Theta}) - \sigma(\tilde{Z}_{\eta(s)}^{1,N,S}, V) \right)^* \, ds \right].
\]

Let us now introduce a parenthesis and notice that since for any given \(a, b \in \mathbb{R}\) we have \(|a^2 - b^2| \leq 2 \max(|a|, |b|) |a - b|\) and since \(\sigma\) is assumed to satisfy the growth condition \((4.1.9)\) uniformly in the external measure component, then:

\[
\mathbb{E} \left[ \left\| \sigma \sigma^*(\tilde{Z}_{\eta(s)}^{1,N,S}, V_{\eta(s)}^{\Theta}) - \sigma \sigma^*(\tilde{Z}_{\eta(s)}^{1,N,S}, V) \right\|_{\mathcal{F}_{\eta(s)}} \right] \\
\leq \mathbb{E} \left[ \left\| (\sigma(\tilde{Z}_{\eta(s)}^{1,N,S}, V_{\eta(s)}^{\Theta}) - \sigma(\tilde{Z}_{\eta(s)}^{1,N,S}, V)) \left( \sigma^*(\tilde{Z}_{\eta(s)}^{1,N,S}, V_{\eta(s)}^{\Theta}) + \sigma^*(\tilde{Z}_{\eta(s)}^{1,N,S}, V) \right) \right\|_{\mathcal{F}_{\eta(s)}} \right] \\
\leq C \mathbb{E} \left[ \left\| \sigma(\tilde{Z}_{\eta(s)}^{1,N,S}, V_{\eta(s)}^{\Theta}) - \sigma(\tilde{Z}_{\eta(s)}^{1,N,S}, V) \right\|_{\mathcal{G}_{\eta(s)}} \left( 1 + |\tilde{Z}_{\eta(s)}^{1,N,S}| \right) \right].
\]

Back to the term \(A_1 := (4.4.14)\), due to the control on the variance of the subsampling, \(A_1 = \mathbb{E} \left[ \int_0^T \frac{1}{2} \text{tr} \left( \sigma(\tilde{Z}_{\eta(s)}^{1,N,S}, V_{\eta(s)}^{\Theta}) (\partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S}) - \partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S})) \sigma^*(\tilde{Z}_{\eta(s)}^{1,N,S}, V_{\eta(s)}^{\Theta}) \right. \\
\left. - \sigma(\tilde{Z}_{\eta(s)}^{1,N,S}, V)(\partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S}) - \partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S})) \sigma^*(\tilde{Z}_{\eta(s)}^{1,N,S}, V_{\eta(s)}^{\Theta}) \right) \, ds \right] \\
\leq C \int_0^T \mathbb{E} \left[ \left\| \partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S}) - \partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S}) \right\|_{\mathcal{G}_{\eta(s)}} \right. \\
\left. \mathbb{E} \left[ \left\| \sigma \sigma^*(\tilde{Z}_{\eta(s)}^{1,N,S}, V_{\eta(s)}^{\Theta}) - \sigma \sigma^*(\tilde{Z}_{\eta(s)}^{1,N,S}, V) \right\|_{\mathcal{G}_{\eta(s)}} \right] \, ds \right] \\
\leq C \int_0^T \mathbb{E} \left[ \left\| \partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S}) - \partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S}) \right\|_{\mathcal{G}_{\eta(s)}} \right. \\
\left. \times \mathbb{E} \left[ \left\| \sigma \sigma^*(\tilde{Z}_{\eta(s)}^{1,N,S}, V_{\eta(s)}^{\Theta}) - \sigma \sigma^*(\tilde{Z}_{\eta(s)}^{1,N,S}, V) \right\|_{\mathcal{G}_{\eta(s)}} \right] \, ds, \right.
\]

which after applying Cauchy–Schwarz inequality, Jensen's inequality for conditional expecta-
tion and the bound we obtained in (4.4.16), satisfies the following:

\[
\leq C \int_0^T \left( \mathbb{E} \left[ \| \partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S}) - \partial_x^2 V(s, \tilde{Z}_{T-\eta(s)}^{1,N,S}) \|^2_{HS} \right] \right)^{1/2} \times \left( \mathbb{E} \left[ \left( 1 + |\tilde{Z}_{\eta(s)}^{1,N,S}| \right) \mathbb{E} \left[ \| \sigma \left( \tilde{Z}_{\eta(s)}^{1,N,S}, v^{\Theta_\eta(s)}, N' \right) - \sigma \left( \tilde{Z}_{\eta(s)}^{1,N,S}, v \right) \|_{HS} \right] \right) \right)^{1/2} \, ds
\]

\[
\leq C \int_0^T \left( \mathbb{E} \left[ \| \partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S}) - \partial_x^2 V(s, \tilde{Z}_{T-\eta(s)}^{1,N,S}) \|^2_{HS} \right] \right)^{1/2} \times \left( \mathbb{E} \left[ \left( 1 + |\tilde{Z}_{\eta(s)}^{1,N,S}| \right)^2 \mathbb{E} \left[ \| \sigma \left( \tilde{Z}_{\eta(s)}^{1,N,S}, v^{\Theta_\eta(s)}, N' \right) - \sigma \left( \tilde{Z}_{\eta(s)}^{1,N,S}, v \right) \|_{HS} \right] \right)^{1/2} \, ds
\]

\[
\leq C \mathcal{R}(S, S')^{1/2} \int_0^T \left( \mathbb{E} \left[ \| \partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S}) - \partial_x^2 V(s, \tilde{Z}_{T-\eta(s)}^{1,N,S}) \|^2_{HS} \right] \right)^{1/2} \, ds.
\]

This and the analogous of the chain of inequalities ending on (4.4.13) but for the next space derivative this time (notice that we will need the fourth space derivative of \( V \) decaying and boundedness of an extra order derivative of the coefficients), namely:

\[
\mathbb{E} \left[ \| \partial_x^2 V(s, \tilde{Z}_{T-s}^{1,N,S}) - \partial_x^2 V(s, \tilde{Z}_{T-\eta(s)}^{1,N,S}) \|^2_{HS} \right] \leq C(h^2 + h) e^{-c\eta(s)} \mathcal{R}(S, S'),
\]

imply that

\[
A_1 \leq Ch^{1/2} \mathcal{R}(S, S').
\]

In the same fashion to the drift, the term \( A_2 := (4.4.15) \) can be treated, due to the property of the diffusion coefficient proved in (4.4.16), Assumption 4.1.4 and uniform moment estimates of at least order \( m/2 + 1 \), as follows:

\[
A_2 = \left[ \mathbb{E} \left[ \int_0^T \frac{1}{2} \text{tr} \left( \sigma \left( \tilde{Z}_{\eta(s)}^{1,N,S}, v^{\Theta_\eta(s)}, N' \right) \partial_x^2 V(s, \tilde{Z}_{T-\eta(s)}^{1,N,S}) \sigma^* \left( \tilde{Z}_{\eta(s)}^{1,N,S}, v^{\Theta_\eta(s)}, N' \right) \right) \right] \right]
\]

\[
\leq C \mathbb{E} \left[ \int_0^T \| \partial_x^2 V(s, \tilde{Z}_{T-\eta(s)}^{1,N,S}) \|^2_{HS} \right] \times \mathbb{E} \left| \tilde{Z}_{\eta(s)}^{1,N,S} \right| \left[ \| \sigma \left( \tilde{Z}_{\eta(s)}^{1,N,S}, v^{\Theta_\eta(s)}, N' \right) - \sigma \left( \tilde{Z}_{\eta(s)}^{1,N,S}, v \right) \|_{HS} \tilde{Z}_{\eta(s)}^{1,N,S} \right] \, ds
\]

\[
\leq C \mathbb{E} \left[ \int_0^T \| \partial_x^2 V(s, \tilde{Z}_{T-\eta(s)}^{1,N,S}) \|^2_{HS} \right] \times \mathbb{E} \left| \tilde{Z}_{\eta(s)}^{1,N,S} \right| \left[ \| \sigma \left( \tilde{Z}_{\eta(s)}^{1,N,S}, v^{\Theta_\eta(s)} \right) - \sigma \left( \tilde{Z}_{\eta(s)}^{1,N,S}, v \right) \|_{HS} \left( 1 + |\tilde{Z}_{\eta(s)}^{1,N,S}| \right) \right] \, ds
\]

\[
\leq C \mathcal{R}(S, S') \int_0^T \mathbb{E} \left[ e^{-cS} \left( 1 + |\tilde{Z}_{T-\eta(s)}^{1,N,S}| \right)^{m/2} \left( 1 + |\tilde{Z}_{\eta(s)}^{1,N,S}| \right) \right] \, ds
\]

\[
\leq C \mathcal{R}(S, S'),
\]

where again \( C \) is a constant independent of \( T, N, N', S \) or \( S' \).

Notice lastly that the term noted above by (4.4.7) is covered, on one hand, by decay of \( \partial_x V \) (Proposition 4.3.10) together with the Lipschitz Assumption 4.1.1 on the space component of
the drift and, on the other hand, as part of Proposition C.5.1 since we need to compare $b$’s action over the empirical measure of the full particle system for $(\bar{Z}_{1,s}^{1,N})_{s \geq 0}$ with its action on the law of the original process $X$. This again together with uniform moment estimates allow us to conclude the uniform in time bound (4.4.7) $\leq C(1 + 1/N)$.

Moreover, since the diffusion is not McKean, i.e. it does not depend on the law of the process itself, the term (4.4.9) is the same as the one treated in Lemma C.4.1 although applied to the process $(\bar{Z}_{1,s}^{1,N})_{s \geq 0}$.

Finally, it is proven that the weak error has order $O(h + 1/N + h^{1/2} \mathcal{R}(S,S') + \tilde{\mathcal{R}}(S,S'))$.

\[ \square \]

4.5 Further research

We conducted all the analysis of the subsampling weak error by usual PDE theory. However, an even more general result is to be expected by using calculus on measure spaces. Indeed, for a non-linear in measure drift and given an $\mathbb{R}$-valued function $U$ on measure spaces, we can proceed analogously using the PDE on measure spaces satisfied by the associated semigroup (known as Master Equation) to study weak errors of the form $[U(\mathcal{L}(X_T^{1})) - \mathbb{E}[U(\mathcal{L}(\bar{Z}_{T}^{1,N}))]]$. The only difference is that, instead of the bounds obtained in Section 4.3, we need decaying in time bounds of the second order linear measure derivatives of $U$, which is not currently available in the literature. However, as mentioned in the introduction, some progress in this field was made recently by [60] on the context of processes on the torus and for particular shapes of non-linearity in the sense of McKean.

Alternatively, decaying bounds for the solution to the Master Equation can be checked in a case by case scenario and it is indeed satisfied for examples such as $U \in C^2_b(\mathbb{R}^d)$ and drifts of the following shape (the associated particle dynamics sometimes referred to as pairwise interactions):

\[ b(x,\mu) = -\beta \partial_x W(x) + \int K(x-y) d\mu(y), \]

for some constant $\beta > 0$, some well behaved potential $\partial_x W$ and binary kernel $K$ (see [83]). In this case, it can be checked computationally that the simulation time is lowered to even more than $1/S$ of the original time just by subsampling.

Special mention deserves the possibility of applying random batch methods to MLMC (with $l$ layers), which is expected to accentuate the benefit from reducing the computational cost from $N^{2l}$ to the much smaller $(SN)^l$ due to the reduction being done on each layer.
Appendices
Appendix A

Appendix Chapter 2

A.1 Non–homogeneous Chapman–Kolmogorov identity

Consider a stochastic process defined for $s \geq 0$ by:

$$X_{T+s}^{r,x} = x + \int_{T}^{T+s} \beta(r, X_{r}^{T,x}) dr + \int_{T}^{T+s} \Sigma(r, X_{r}^{T,x}) dB_r,$$  \hspace{1cm} (A.1.1)

and which has a density $[\tau, \infty) \times \mathbb{R}^d \ni (\tau+s, x') \rightarrow p_{\tau+s}^{T}(x, x')$. Then, for any function $\phi \in C(\mathbb{R}^d)$ with compact support, we have that

$$E[\phi(X_{\tau+s}^{T,x})] = \int_{\mathbb{R}^d} \phi(x') p_{\tau+s}^{T}(x, x') dx'.$$

Lemma A.1.1. Assume the process $Y$ defined by (A.1.1) satisfies the flow property [78, Theorem 3.3]. Then, its density $[\tau, \infty) \times \mathbb{R}^d \ni (\tau+s, x') \rightarrow p_{\tau+s}^{T}(x, x')$ satisfies the following for any $\tau \leq 1$, $s \geq 0$ and $x, x'' \in \mathbb{R}^d$:

$$p_{\tau+s}^{T}(x, x'') = \int p_{\tau+s}^{1}(x', x'') p_{\tau}^{1}(x, x') dx'.$$  \hspace{1cm} (A.1.2)

Proof. For $\tau \leq 1$ recall that we assumed for any $\tau, s, x, X_{\tau+s}^{T,x} = X_{\tau+s}^{1+\tau, X_{\tau+s}^{T,x}}$.

First by the tower property and afterwards by the flow property above, for any $\phi \in C(\mathbb{R}^d)$ with compact support we have:

$$\int_{\mathbb{R}^d} \phi(x') p_{\tau+s}^{T}(x, x') dx' = E[\phi(X_{\tau+s}^{T,x})] = E[E[\phi(X_{\tau+s}^{1, X_{\tau+s}^{T,x}})]| X_{\tau}^{T,x}]$$

$$= \int_{\mathbb{R}^d} E[\phi(X_{\tau+s}^{1, x''})] p_{\tau}^{1}(x, x'') dx'' = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \phi(x') p_{\tau+s}^{1}(x'', x') p_{1}^{T}(x, x'') dx'' dx'$$

$$= \int_{\mathbb{R}^d} \phi(x') \left( \int_{\mathbb{R}^d} p_{\tau+s}^{1}(x'', x') p_{1}^{T}(x, x'') dx'' \right) dx'.$$

Finally, since the above holds for an arbitrary test function $\phi$, we conclude (A.1.2). \qed
Appendix B

Appendix Chapter 3

B.1 Moments estimates

Lemma B.1.1. Suppose that Assumption 3.2.3 holds and let us name $\tilde{M}_1 = M_1 / 2$ and recall that $M_1 > 2M_0 > 0$. When $\sigma$ is bounded the following holds for some constant $M_2 \in \mathbb{R}$ and any $2 \leq m \leq p$:

$$
\langle x, b(x, \mu) \rangle + \frac{(m-1)}{2} |\sigma(x, \mu)|^2 \leq M_2 - \tilde{M}_1 |x|^2 + M_0 \int_{\mathbb{R}^d} |x|^2 d\mu(x), \quad \forall (x, \mu) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d).
$$

(B.1.1)

Moreover, if $\sigma$ is not bounded, for any $\epsilon > 0$ there exist some other constant $\tilde{M}_2 \in \mathbb{R}$ such that

$$
\langle x, b(x, \mu) \rangle + \frac{(m-1-\epsilon)}{2} |\sigma(x, \mu)|^2 \leq \tilde{M}_2 - \tilde{M}_1 |x|^2 + M_0 \int_{\mathbb{R}^d} |x|^2 d\mu(x), \quad \forall (x, \mu) \in \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d).
$$

(B.1.2)

Proof. When choosing $(x_0, \mu_0) = (0, \delta_0)$ and for any $(x, \mu)$, the monotonicity Assumption 3.2.3 implies by the use of the Triangle and Young’s inequalities that

$$
\langle x, b(x, \mu) \rangle + \frac{(m-1)}{2} |\sigma(x, \mu)|^2 - (m-1)|\sigma(x, \mu)\sigma^*(0, \delta_0)| \leq \frac{(m-1)}{2} |\sigma(0, \delta_0)|^2
$$

$$
= \langle x - 0, b(x, \mu) - b(0, \delta_0) \rangle + \langle x - 0, b(0, \delta_0) \rangle + \frac{m-1}{2} |\sigma(x, \mu) - \sigma(0, \delta_0)|^2
$$

$$
\leq \langle x, b(0, \delta_0) \rangle - M_1 |x|^2 + M_0 \int_{\mathbb{R}^d} |x|^2 d\mu(x)
$$

$$
\leq \frac{1}{2M_1} |b(0, \delta_0)|^2 - \frac{M_1}{2} |x|^2 + M_0 \int_{\mathbb{R}^d} |x|^2 d\mu(x).
$$

(B.1.3)

Moreover, if $\sigma$ is assumed to be bounded, we have that

$$
\langle x, b(x, \mu) \rangle + \frac{(m-1)}{2} |\sigma(x, \mu)|^2
$$

$$
\leq \frac{1}{2M_1} |b(0, \delta_0)|^2 + \frac{m-1}{2} |\sigma(0, \delta_0)|^2 + \frac{m-1}{2} |\sigma||^2_{\infty} - \frac{M_1}{2} |x|^2 + M_0 \int_{\mathbb{R}^d} |x|^2 d\mu(x).
$$

Hence, there exist some $M_2 := \frac{1}{M_1} |b(0, \delta_0)|^2 + \frac{m-1}{2} |\sigma(0, \delta_0)|^2 + \frac{m-1}{2} |\sigma||^2_{\infty} \in \mathbb{R}$ such that if $\tilde{M}_1 := M_1 / 2$, (B.1.1) is satisfied and $\tilde{M}_1 > M_0 > 0$. 

128
However, if $\sigma$ is not bounded, we can instead iterate the use of Young’s inequality in its most general shape for some $\epsilon > 0$ and obtain from (B.1.3) that

$$\langle x, b(x, \mu) \rangle + \frac{(m-1)}{2} |\sigma(x, \mu)|^2$$

$$\leq \frac{1}{2M_1} |b(0, \delta_0)|^2 + (m-1) \left( \frac{\epsilon}{2(m-1)} |\sigma(x, \mu)|^2 + \frac{(m-1)}{2\epsilon} |\sigma(0, \delta_0)|^2 \right)$$

$$- \frac{M_1}{2} |x|^2 + M_0 \int_{\mathbb{R}^d} |x|^2 d\mu(x).$$

and finally, for $m \geq 2$,

$$\langle x, b(x, \mu) \rangle + \frac{(m-1-\epsilon)}{2} |\sigma(x, \mu)|^2$$

$$\leq \frac{1}{M_1} |b(0, \delta_0)|^2 + \frac{(m-1)^2}{2\epsilon} |\sigma(0, \delta_0)|^2 - \frac{M_1}{2} |x|^2 + M_0 \int_{\mathbb{R}^d} |x|^2 d\mu(x),$$

i.e. there exist some $\tilde{M}_2 := \frac{1}{M_1} |b(0, \delta_0)|^2 + \frac{(m-1)^2}{2\epsilon} |\sigma(0, \delta_0)|^2 \in \mathbb{R}$ such that (B.1.2) for $\tilde{M}_1 = \bar{M}_1/2 > M_0 > 0$, slightly less tight than (B.1.1), is satisfied.

For completeness, we present next the result which can also be found in Lemma C.2.5, but with the notation from Chapter 3.

**Lemma B.1.2.** Given $\xi \in L^m(\mathbb{R}^d)$, let Assumptions 3.2.1, 3.2.3 hold with $p > 2m$. Then, the process $(X_s^\xi)_{s \geq 0}$ defined by (3.1.1) satisfies the following:

$$\mathbb{E}[|X_s^\xi|^2] \leq \frac{M_2}{M_1 - M_0} + e^{-2(\bar{M}_1 - M_0)s} \mathbb{E}[|\xi|^2]$$

and for $m > 2$,

$$\mathbb{E}[|X_s^\xi|^m] \leq \frac{M_2}{(M_1)^2} + \frac{m}{(m-2)M_1} e^{-2(\bar{M}_1 - M_0)s} \mathbb{E}[|\xi|^2] + e^{-m\bar{M}_1 s} \mathbb{E}[|\xi|^m],$$

where $M_0, \bar{M}_1, M_2$ are the constants from Lemma (B.1.1). In particular, there exist constants $C, c > 0$ such that:

$$\mathbb{E}[|X_s^\xi|^m] \leq C(1 + e^{-mcse}[|\xi|^m]).$$

**Proof.** Let $\lambda$ be an arbitrary positive constant and $\xi$ a starting random variables. If Itô’s formula is applied to the process $(e^{\lambda s}X_s^\xi |_{s \geq 0})$, we obtain:

$$d(e^{\lambda s}X_s^\xi |_{s \geq 0}) = e^{\lambda s} \left( \lambda |X_s^\xi|^m + m|X_s^\xi|^{m-2} \left( X_s^\xi, b(X_s^\xi, \mathcal{L}(X_s^\xi)) \right) 
\right. 
+ \frac{m(m-1)}{2} |X_s^\xi|^{m-2} \operatorname{tr} \left( \sigma \sigma^* (X_s^\xi, \mathcal{L}(X_s^\xi)) \right) ds 
\left. + e^{\lambda s} \left( m|X_s^\xi|^{m-2} \left( X_s^\xi, \sigma(X_s^\xi, \mathcal{L}(X_s^\xi)) dB_s \right) \right). \right.$$
which means that for the choice of \( \lambda \) which can be treated with the coercivity condition (B.1.2) in order to obtain

\[
e^{\lambda s} \left( (\lambda - \bar{M}_1) m \mathbb{E}[|X_s^\xi|^m] + m M_0 \mathbb{E}[|X_s^\xi|^2 + m M_2] \right) ds,
\]

which can be treated with the coercivity condition (B.1.2) in order to obtain

\[
\leq e^{\lambda s} \left( (\lambda - \bar{M}_1) m \mathbb{E}[|X_s^\xi|^m] + m M_0 \mathbb{E}[|X_s^\xi|^2 + m M_2] \right) ds.
\]

Notice next that in the case of \( m = 2 \), we obtain:

\[
d\mathbb{E}[e^{\lambda s}|X_s^\xi|^2] \leq e^{\lambda s} \left( (\lambda - 2 \bar{M}_1) \mathbb{E}[|X_s^\xi|^2] + 2 M_0 \mathbb{E}[|X_s^\xi|^2 + 2 M_2] \right) ds,
\]

which means that for the choice of \( \lambda = 2(\bar{M}_1 - M_0) > 0 \),

\[
d\mathbb{E}[e^{2(\bar{M}_1 - M_0)s}|X_s^\xi|^2] \leq 2 M_2 e^{2(\bar{M}_1 - M_0)s} ds.
\]

After integration, we conclude that

\[
\mathbb{E}[|X_s^\xi|^2] \leq \frac{M_2}{\bar{M}_1 - M_0} + e^{-2(\bar{M}_1 - M_0)s} \mathbb{E}[|\xi|^2].
\]

Going back to (B.1.4), we observe that

\[
d\mathbb{E}[e^{\lambda s}|X_s^\xi|^m] \leq e^{\lambda s} \left( (\lambda - \bar{M}_1) m \mathbb{E}[|X_s^\xi|^m] + m \left( \frac{M_2}{\bar{M}_1 - M_0} + e^{-2(\bar{M}_1 - M_0)s} \mathbb{E}[|\xi|^2] \right) \right) ds.
\]

Hence, we can make the particular choice \( \lambda = m \bar{M}_1 > 0 \), (B.1.5) becomes

\[
d\mathbb{E}[e^{m \bar{M}_1 s}|X_s^\xi|^m] \leq m \left( \frac{M_2}{\bar{M}_1 - M_0} + e^{-2(\bar{M}_1 - M_0)s} \mathbb{E}[|\xi|^2] \right) e^{m \bar{M}_1 s} ds,
\]

and by integrating in between 0 and \( s \) and using Hölder’s inequality,

\[
\mathbb{E}[|X_s^\xi|^m] \leq \frac{M_2}{\bar{M}_1 (\bar{M}_1 - M_0)} + \frac{m}{m \bar{M}_1 - 2 (\bar{M}_1 - M_0)} e^{-2(\bar{M}_1 - M_0)s} \mathbb{E}[|\xi|^2] + e^{-m \bar{M}_1 s} \mathbb{E}[|\xi|^m]
\]

\[
\leq \frac{M_2}{(\bar{M}_1)^2} + \frac{m}{(m - 2) \bar{M}_1} e^{-2(\bar{M}_1 - M_0)s} \mathbb{E}[|\xi|^2] + e^{-m \bar{M}_1 s} \mathbb{E}[|\xi|^m].
\]

In other words, uniform in time bounds of the moments.

In particular, for \( C = \max \{ M_2/(\bar{M}_1)^2, m/(m - 2) \bar{M}_1 \} \) and \( c = \min \{ 2(\bar{M}_1 - M_0)/m, \bar{M}_1 \} \),

\[
\mathbb{E}[|X_s^\xi|^m] \leq C M_2 + e^{-c m s} \mathbb{E}[|\xi|^2] + |\xi|^m.
\]

\[\square\]

**Lemma B.1.3.** Given \( \xi, \psi \in L^m(\mathbb{R}^d) \), let Assumptions 3.2.1, 3.2.3 hold with \( p > 2m \). Then, the process \( (Y_s^\tau, \psi; \xi)_{s \geq \tau} \) defined by (3.1.1) satisfies the following moment bounds. There exist
constants $C, c_1, c_2 > 0$ such that:

$$\mathbb{E}[|Y_{s}^{T,\psi;\xi}|^m] \leq C(1 + e^{-(m/2)(s-\tau)}\mathbb{E}[|\psi|^m] + e^{-c_2(s-\tau)}\mathbb{E}[|\xi|^m]).$$

Notice that any lower order moment will have the same type of uniform bound by the same reasoning.

**Proof.** Let $\lambda$ be an arbitrary positive constant and $\psi, \xi \in L_m^m(\mathbb{R}^d)$ starting random variables. If Itô’s formula is applied to the process $(e^{\lambda s} Y_s^{T,\psi;\xi})_{s \geq \tau}$, we obtain:

$$d\left(e^{\lambda s} Y_s^{T,\psi;\xi}\right) = e^{\lambda s} \left(\lambda Y_s^{T,\psi;\xi} + m Y_s^{T,\psi;\xi} - (Y_s^{T,\psi;\xi}, \mathcal{L}(X_s^\xi))\right) + \frac{m(m-1)}{2} Y_s^{T,\psi;\xi} \text{tr}(\sigma^*(Y_s^{T,\psi;\xi}, \mathcal{L}(X_s^\xi))) ds + \left(e^{\lambda s} (m Y_s^{T,\psi;\xi} - (Y_s^{T,\psi;\xi}, \mathcal{L}(X_s^\xi))) dB_s\right).$$

And now, by taking expectations of the above and after a stopping time argument, we arrive to

$$d\mathbb{E}[e^{\lambda s} Y_s^{T,\psi;\xi}] = e^{\lambda s} \left(\lambda \mathbb{E}[Y_s^{T,\psi;\xi}] + m \mathbb{E}[Y_s^{T,\psi;\xi}] - (Y_s^{T,\psi;\xi}, \mathcal{L}(X_s^\xi))\right) + \frac{m(m-1)}{2} \mathbb{E}[Y_s^{T,\psi;\xi}] \text{tr}(\sigma^*(Y_s^{T,\psi;\xi}, \mathcal{L}(X_s^\xi))) ds,$$

which can be treated with the coercivity condition (B.1.2) and bounds in Lemma B.1.2 in order to obtain

$$\leq e^{\lambda s} \left((\lambda - \bar{M}_1 m)\mathbb{E}[Y_s^{T,\psi;\xi}] + \mathbb{E}[Y_s^{0,\psi;\xi}] - (Y_s^{0,\psi;\xi}, \mathcal{L}(X_s^\xi))\right) ds$$

$$\leq e^{\lambda s} \left((\lambda - \bar{M}_1 m)\mathbb{E}[Y_s^{T,\psi;\xi}] + \mathbb{E}[Y_s^{0,\psi;\xi}] - (Y_s^{0,\psi;\xi}, \mathcal{L}(X_s^\xi))\right) ds.$$

(B.1.6)

Next we use Young’s inequality (namely for any $\epsilon > 0$, $a = m M_2 + \frac{m M_0 M_2}{\bar{M}_1 - M_0} + m M_0 e^{-2(M_1 - M_0) s} \mathbb{E}[|\xi|^2]$), $b = |Y_s^{T,\psi;\xi}|$, $p = m/2$, $q = m/(m-2)$ we have $ab \leq \frac{a^p}{2\epsilon^{p/q}} + \frac{b^q}{2}$:

$$d\mathbb{E}[e^{\lambda s} Y_s^{T,\psi;\xi}] \leq e^{\lambda s} \left((\lambda - \bar{M}_1 m + \epsilon/2)\mathbb{E}[Y_s^{0,\psi;\xi}] + \frac{a^p}{2\epsilon^{p/q}}\right) ds.$$

Hence, if we fix $\epsilon$ small enough to guarantee that we can make the particular choice $\lambda = m \bar{M}_1 - \epsilon/2 > 0$, it becomes, for some $C, c_2 > 0$,

$$d\mathbb{E}[e^{(m \bar{M}_1 - \epsilon/2) s} Y_s^{T,\psi;\xi}] \leq C \left(1 + e^{-c_2 s} \mathbb{E}[|\xi|^m]\right)^{e^{(m \bar{M}_1 - \epsilon/2) s}} ds,$$

and by integrating in between $\tau$ and $s$ the result is

$$\mathbb{E}[|Y_s^{T,\psi;\xi}|^m] \leq C \left(1 + e^{-(m \bar{M}_1 - \epsilon/2)(s-\tau)}\mathbb{E}[|\psi|^m] + e^{-c(s-\tau)}\mathbb{E}[|\xi|^m]\right),$$

in other words, the uniform in time moment bounds stated with $c_1 = \bar{M}_1 - \epsilon/2m$. \qed
Remark B.1.4. Let Assumption 3.2.1 hold. Then, for given any \( x, x_0, y, y_0 \in \mathbb{R}^d \) and once we set \( \mu = \mathcal{L}(X_t^x) \) and \( \mu_0 = \mathcal{L}(X_t^{x_0}) \),

\[
\langle y - y_0, b(y, \mathcal{L}(X_t^x)) - b(y_0, \mathcal{L}(X_t^{x_0})) \rangle \leq -M_1 |y - y_0|^2 + M_0 \mathbb{E}[|X_t^x - X_t^{x_0}|^2].
\]

Consequently on one hand, by the bounds obtained in Lemma 3.2.10, there exists \( C > 0 \) such that for all \( t \geq 0 \),

\[
\langle y - y_0, b(y, \mathcal{L}(X_t^x)) - b(y_0, \mathcal{L}(X_t^{x_0})) \rangle = \langle y - y_0, b(y, \mathcal{L}(X_t^x)) - b(y_0, \mathcal{L}(X_t^{x_0})) \rangle \\
\leq C|x - x_0|^2.
\]

And on the other hand, by taking \( x = x_0 \in \mathbb{R}^d \) and any \( y, y_0, t \),

\[
\langle y - y_0, b(y, \mathcal{L}(X_t^x)) - b(y_0, \mathcal{L}(X_t^{x_0})) \rangle \leq -M_1 |y - y_0|^2.
\]

Both together imply that any \( x, x_0, y, y_0, t \)

\[
\langle y - y_0, b(y, \mathcal{L}(X_t^x)) - b(y_0, \mathcal{L}(X_t^{x_0})) \rangle \leq C|x - x_0|^2 + M_1 |y - y_0|^2.
\]

Let us fix next an orthonormal basis of \( \mathbb{R}^d : \{ e_1, ..., e_d \} \), \( h > 0 \) and two indexes \( i, j \in \{ 1, ..., d \} \). Now, \( y_0 \) was arbitrary so it can be taken so that \( y - y_0 = he_j \) and conclude that:

\[
\frac{\langle y - y_0, b_j(y, \mathcal{L}(X_t^x)) - b_j(y, \mathcal{L}(X_t^{x_0} + he_j)) \rangle}{h|y - y_0|} = \lim_{h \to 0} \frac{b_j(y, \mathcal{L}(X_t^x)) - b_j(y, \mathcal{L}(X_t^{x_0} + he_j))}{h^2} \\
\leq \lim_{h \to 0} \frac{C h^2 |e_j|^2 + M_1 h^2 |e_j|^2}{h^2},
\]

or in other words: if the derivative \( \partial_x b((y, \mathcal{L}(X_t^x))) \) exists.

Remark B.1.5. Fixed a parameter \( x \in \mathbb{R}^d, (s, z) \rightarrow \partial_x \beta(s, \cdot ; x)z + \partial_z \Sigma(s, \cdot ; x)z \) is coercive whenever \( \beta, \Sigma \) are obtained from \( b, \sigma \) as above display (4.3.1) and these satisfy the monotonicity Assumption 3.2.3.

Indeed, because of [91, Preposition 2.20], we know that for any fixed \( j \in \{ 1, ..., d \} \)

\[
\sup_{\mathbb{R}^d z \neq 0} \frac{\langle z, ((z, \partial_y \beta_j(s, y; x))) \rangle + m^{-1} \|((z, \partial_y \Sigma^{j=1}(s, y; x))) \|_{HS}^2}{\|z\|^2} \\
\leq \sup_{y_1, y_2 \in \mathbb{R}^d} \frac{\langle y_1 - y_2, \beta(s, y_1; x) - \beta(s, y_2; x) + m^{-1} \|\Sigma(s, y_1; x) - \Sigma(s, y_2; x)\|_{HS}^2}{\|y_1 - y_2\|^2},
\]

where we used the notation \( \| \cdot \|_{HS} \) for the Hilbert-Schmidt norm of matrices. From this together with Proposition 3.2.10, we have that when \( Z_s \neq 0 \) the following inequality holds pathwise for any \( x \in \mathbb{R}^d \):

\[
\frac{\langle Z_s^i, ((Z_s^i, \partial_y \beta_j(s, Y_s^{y; x})) \rangle + m^{-1} \|((Z_s^i, \partial_y \Sigma^{j=1}(s, Y_s^{y; x})) \|_{HS}^2}{\|Z_s^i\|^2}
\]

132
We next recall the analogous result to Lemma 4.3.9.

**Lemma B.1.6.** Under Assumptions 3.2.1, 3.2.2 and 3.2.3, the process \((Z_s)_{s \geq 0}\) defined by (3.2.6) satisfies the following bounds for some constant \(C > 0\) and any \(4 \leq 2m < p, \chi, \xi \in L^m(\mathbb{R}^d)\):

\[
\mathbb{E}[|Z_s^{\chi, \xi}|^m] \leq Ce^{-sM_1}, \quad \text{for all } s \geq 0.
\]

**B.2 Examples with decaying bounds for \(\partial^2_j V\)**

**B.2.1 Lyapunov assumptions**

Given \(K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d\), \(b : \mathbb{R}^d \to \mathbb{R}^d\) and \(\xi \in L^2(\mathbb{R}^d)\), consider the following SDE for some parameter \(\epsilon \in \mathbb{R}\) fixed:

\[
dX_s^\xi = \epsilon \int K(X_s^\xi, y) \mathcal{L}(X_s^\xi)(dy) + b(X_s^\xi) ds + dB_s, \quad \forall s \in [0, \infty); \quad X_0^\xi = \xi. \tag{B.2.1}
\]

**Assumption B.2.1.** [Regularity and growth coefficients] We assume \(K, b\) have bounded first and second derivatives and there exist \(L, \kappa > 0\) such that

\[
|K(x, x') - K(y, y')| \leq L(|x - x'| + |y - y'|), \quad \text{for any } x, x', y, y' \in \mathbb{R}^d;
\]

\[
|b(x) - b(x')| \leq \kappa |x - x'|, \quad \text{for any } x, x' \in \mathbb{R}^d.
\]

**Assumption B.2.2.** [Well posedness and existence of invariant measure]

There exists a function \(W \in C^2(\mathbb{R}^d)\) such that \(\inf_{x \in \mathbb{R}^d} W(x) > 0\). Moreover,

1. there exists a strictly positive, increasing and concave \(C^1(\mathbb{R}_+)\) function \(\eta\) such that

\[
\lim_{|x| \to \infty} \eta(W(x)) = \infty \quad \text{and a constant } C > 0 \text{ satisfying:}
\]

\[
LW(x) \leq C - \eta(W(x)) \quad \text{for any } x \in \mathbb{R}^d;
\]

2. there exists a constant \(\alpha > 0\) and a bounded set

\(S_2 \supset S_1 := \{(x, y) \in \mathbb{R}^d \times \mathbb{R}^d : \eta(W(x)) + \eta(W(y)) \leq 4C\}\) such that for any \(x, y \in \mathbb{R}^d \setminus S_2\)

\[
\eta(W(x)) + \eta(W(y)) \geq \max\{4C, 1/\int_0^{R_1} \psi(r)^{-1} dr, 1 + \alpha \int_0^{R_1} \psi(r)^{-1} d r \eta(\Phi(|x - y|))\},
\]

where \(R_1 := \sup\{|x - y| : (x, y) \in S_1\}\), \(\psi(r) = \exp(-1/4\kappa r^2)\) and \(\Phi(r) = \int_0^r \psi(t) dt\).

Plenty of examples satisfying this set of assumptions can be found in [33].
Finally, we consider test functions as the set of bounded, measurable functions on \( \mathbb{R}^d \), bounded by a fixed constant \( M \), i.e.

\[
\mathcal{S}_M := \left\{ f \in B(\mathbb{R}^d) \text{ s.t. } \sup_{x \in \mathbb{R}^d} |f(x)| \leq M \right\}.
\]

Again, we construct the auxiliary process (4.3.1), where this time

\[
\beta(t, y; x) = \epsilon \int K(y, z) \mathcal{L}(X_s^t)(dz) + b(y) \quad \text{and} \quad \Sigma(t, y; x) = I_d,
\]

in order to define the function \((s, y, x) \mapsto V(s, y; x) = \mathbb{E}[\phi(\gamma_{s,x})] \) for an arbitrary element \( \phi \in \mathcal{S} \).

**Theorem B.2.3.** Suppose that Assumptions B.2.2 and B.2.1 hold for the process defined by the SDE (B.2.1) and that there exists \( m \in \mathbb{N}, M > 0 \) such that \( |W(x)| \leq M(1 + |x|^m), \forall x \). Fix \( \phi \in \mathcal{S} \) and \( t \to H(t) = \int_1^t \frac{1}{\eta(s)} ds \). Then, for \( n = 1, 2 \) there exists \( C > 0 \) such that

\[
|\partial_y^n V(s, y; x)|_{y=x} \leq C e^{(v+c)|x|^2} H(cs), \quad \text{for all } (s, x) \in (1, \infty) \times \mathbb{R}^d.
\]

Notice that since \( \eta(s) \) is concave, \( H^{-1} \) maps \([0, \infty)\) to \([1, \infty)\); \( \lim_{s \to \infty} H^{-1}(s) = \infty \), but \( H^{-1} \) might not always be increasing.

**Proof Theorem B.2.3.** Let us see first that the results [26, Theorem VI.2, Theorem VI.5] are valid which, just as in Lemma 2.3.8, provide quantitative bounds for the derivatives of the transition density at time \( t = 1 \). Since the diffusion coefficient is the identity, the uniform ellipticity of the infinitesimal generator is ensured. Additionally, from the uniform Lipschitz conditions we conclude the linear growth of the drift if this is not null everywhere. The boundedness of the derivatives finally complete the check list. Hence,

\[
\mathbb{E}\left[1 + |x'\cdot x''|\right] \int \partial_y^n p_1^0(y, x''; x)|_{y=x} dx'' \leq C e^{v|x|^2} \left( \int \mathbb{E}\left[1 + |x''|^m\right] e^{-v|x''|^2} e^{c|x|^2} |x''|^2 dx'' \right)
\]

meaning that Assumption 3.1.3(i) holds for some \( C, c > 0 \) with \( \mathbb{R}^d \ni x \to g(x) = C(1 + |x|^m) \) and \( \mathbb{R}^d \ni x \to h(x) = C e^{c|x|^2} \). Assumption 3.1.3(ii) follows by the same Lemma 2.3.8.

Next, from [33, Corollary 2.5] (which still holds for locally Lipschitz coefficients \( b \)) we know that for small enough \( \epsilon > 0 \), fixed \( x \in \mathbb{R}^d \) and \( (s, x''') \to P_{1+s}^1(x, x''') \) the density of \( X_{1+s}^x \),

\[
\left\| P_{1+s}^1(x, \cdot) - q \right\|_{TV} \leq \frac{R_2 + \epsilon W(x)}{H^{-1}(cs)} + C \frac{2 \epsilon (\eta^{-1} 2 C l) + 1}{2 C l H^{-1}(cs)}, \quad \text{for any } x \in \mathbb{R}^d \text{ and } s \geq 0,
\]

where \( l = 2 \epsilon \inf_{x \in \mathbb{R}^d} W(x) \) and \( H(t) = \int_1^t \frac{1}{\eta(s)} ds \).

Next, on one side for any \( \phi \in B(\mathbb{R}^d) \) bounded by \( M \in \mathbb{R}_+ \),

\[
\int \mid\phi(x'')\mid P_{1+s}^1(x', dx''') - q(dx'') \mid dx'' \leq M \left( \frac{R_2 + \epsilon W(x')}{H^{-1}(cs)} + C \frac{2 \epsilon (\eta^{-1} 2 C l) + 1}{2 C l H^{-1}(cs)} \right)
\]

\[
\leq CH(cs)(1 + W(x'));
\]

134
and on the other side, for some $c, \nu > 0$,
\[
\int_{\mathbb{R}^d} e^{-c|x-x'|^2 + \nu |x|^2} H(cs)(1 + W(x')) d x' \leq C H(cs) e^{(\nu + c)|x|^2}.
\]
In other words, since $X_s = Y_s^{\tau, x}$ a.s, Assumption 3.1.2 is satisfied with $G(s) = CH(cs), g(x) = C(1 + |x|^m)$ and $h(x) = e^{(\nu + c)|x|^2}$.

Finally, by Theorem 3.1.4, we then conclude that $\partial_y V, \partial^2_y V$ are bounded uniformly in time. More specifically, at any time $s$ they are bounded by $\frac{1}{H^{-1}(cs)}$.

\[\square\]

### B.2.2 Veretennikov–Khasminskii assumptions

Given $\xi \in L^2(\mathbb{R}^d)$, $c > 0$ and a constant diffusion $\sigma > 0$, we consider the following type of SDE:

\[
d X_s^{\tau, \xi} = b_1(X_s^{\tau, \xi}) + c b_2(X_s^{\tau, \xi}, \mathcal{L}(X_s^{\tau, \xi})) d s + \sigma d B_s, \quad \forall s \in [\tau, \infty);
X_\tau^{\tau, \xi} = \xi.
\]  

(B.2.2)

The idea behind this model being that a small, non-linear in the sense on McKean, perturbation of a well behaved usual autonomous SDE with additive noise remains well behaved.

**Assumption B.2.4.** [Veretennikov—Khasminskii–type conditions] For any $x \in \mathbb{R}^d$ there exist some $M, M_1 > 0$ such that:

\[
\langle x, b_1(x) \rangle \leq -M_1|x|, \quad \forall |x| \geq M;
\]

**Assumption B.2.5.** [Regularity and growth coefficients] The coefficients $b_1, b_2$ are Lipschitz: there exists $L > 0$ such that for all $x, y \in \mathbb{R}^d$ and $\mu, v \in \mathcal{P}(\mathbb{R}^d)$,

\[
|b_1(x) - b_1(y)| + |b_2(x, \mu) - b_2(y, v)| 
\leq L \left( |x - y| + \left( \inf_{\pi \in \Pi(\mu, v)} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \min \{1, |x - y|^2\} \pi(dx, dy)^{1/2} \right) \right),
\]

where $\Pi(\mu, v)$ is the family of couplings in between $\mu, v$. Assume moreover that $b_1, b_2$ have bounded derivatives up until second order ($\partial_y$ is understood in the sense of intrinsic derivative). Additionally, $b_2$ is bounded uniformly (i.e $\sup_{x \in \mathbb{R}^d, \mu \in \mathcal{P}(\mathbb{R}^d)} b_2(x, \mu) \leq M, \text{ for some } M > 0$) and $b_1$ is Hölder continuous (i.e there exist $L > 0, \lambda \in (0, 1)$ such that $|b_1(x) - b_1(y)| \leq L|x - y|^\lambda, \forall x, y \in \mathbb{R}^d$).

Finally, fixed $M^* \in \mathbb{R}_+$ we restrict ourselves to the possible evaluation functions $\phi$ within the family

\[
\mathcal{S} := \left\{ f \in B(\mathbb{R}^d) : \sup_{x \in \mathbb{R}^d} |f(x)| \leq M^* \right\}.
\]

Since existence of a unique strong solution to (B.2.2), $(X_s^{\tau, \xi}, \mathcal{L}(X_s^{\tau, \xi}))_{t > 0}$, is well known in the Lipschitz case, once we fix $\xi = x \in \mathbb{R}^d$, we proceed constructing the auxiliary process:

\[
d Y_s^{\tau, y, x} = b_1(Y_s^{\tau, y, x}) + c b_2(Y_s^{\tau, y, x}, \mathcal{L}(X_s^{\tau, x})) d s + \sigma d B_s, \quad \forall s \in [\tau, \infty); \quad Y^{\tau, y, x}_t = y;
\]  

(B.2.3)
whose density we denote by $x'' \rightarrow p^x_s(y, x''; x)$.

**Theorem B.2.6.** Consider $V(s, y; x) := E[\phi(Y^0_s, y, x)]$, where $(Y^0_s, y, x)_{s \geq 0}$ is defined by (B.2.3) and $\phi \in \mathcal{S}$. Suppose that Assumptions B.2.5 and B.2.4 hold. Then, there exists $\epsilon_0 > 0$ such that for any $\epsilon \in [0, \epsilon_0)$ and for $n = 1, 2$ there exists $k > 0$ such that:

$$\left| \partial^n_y V(s, y; x) \right|_{y=x} \leq e^{-k s} C (1 + e^{(\nu + c)|x|^2}), \quad \text{for all } (s, x) \in [1, \infty) \times \mathbb{R}^d.$$

**Proof.** First notice that [92, Theorem 3.1] guarantees exponential convergence in Total Variation for $\epsilon$ below some threshold $\epsilon_0 > 0$. Namely, if we denote by $(s, x'') \rightarrow P^1_s(x, x'')$ the density of $X^{1,x}_{1+s}$, there exist $q = q^x \in \mathcal{P}(\mathbb{R}^d)$ and $C, k > 0$ such that

$$\| p^1_s(y, x'') \|_{y=x} - q(\cdot) \| TV \leq \| P^1_s(x, \cdot) - q(\cdot) \| TV \leq C (1 + e^{1|x|^2}) e^{-k s}.$$

Now, this together with our set of test functions $\mathcal{S}$ implies convergence in Weighted Total Variation, i.e. for any $\phi \in B(\mathbb{R}^d)$ bounded,

$$\left| \int_{\mathbb{R}^d} \phi(x'') (p^1_s(y, d x''); x)_{y=x} - q(d x'') \right| \leq M^* \int_{\mathbb{R}^d} \| p^1_s(y, d x''); x)_{y=x} - q(d x'') \| \leq M^* (1 + e^{1|x|^2}) e^{-k s}.$$

Hence, Assumption 2.2.2 is satisfied for some $C > 0$ with $\mathbb{R}^d \ni x \mapsto g(x) = C (1 + e^{1|x|^2})$ and $(0, \infty) \ni s \mapsto G(s) = e^{-k s}$.

Next, by Assumption B.2.4 and given the specific shape of the SDE at hand, all the conditions in [26, Theorem VI.5] are satisfied and for $n \leq 2$ the following holds:

$$\left| \partial^n_x p^1_s(y, x''; x) \right|_{y=x} \leq C e^{-c|x-x''|^2 + 2|x|^2 - 2|x''|^2}.$$

Therefore, similarly to Theorem 2.3.8, Assumption 3.1.2 holds with the following bounds:

$$\int_{\mathbb{R}^d} (1 + e^{1|x''|^2}) \left| \partial^n_x p^1_s(y, x''; x) \right|_{y=x} d x'' \leq C \int_{\mathbb{R}^d} (1 + e^{1|x''|^2}) e^{-c|x-x''|^2 + 2|x|^2 - 2|x''|^2} d x''$$

$$\leq C \int_{\mathbb{R}^d} (1 + e^{1|x''|^2}) e^{-c|x-x''|^2 + 2|x|^2 - 2|x''|^2} d x''$$

$$\leq C (1 + e^{(c+v)|x|^2}),$$

i.e for some $C > 0$ with $\mathbb{R}^d \ni x \mapsto g(x) = C (1 + e^{1|x|^2})$ and $\mathbb{R}^d \ni x \mapsto h(x) = C e^{(c+v)|x|^2}$. Finally Theorem 3.1.4 can be applied.

\hfill \square

### B.2.3 Another set of Lyapunov conditions

We consider another type of small non-linear perturbation in the sense of McKean: for some $\epsilon > 0$ fixed,

$$d X^\xi_s = \int_{\mathbb{R}^d} K_\epsilon(X^\xi_s, y) \mathcal{L}(X^\xi_s)(dy) + \epsilon b_0(X^\xi_s, \mathcal{L}(X^\xi_s)) ds + \sigma(X^\xi_s) dB_s, \quad \forall s \in [0, \infty);$$

\begin{equation}
X^\xi_0 = \xi. \tag{B.2.4}
\end{equation}
For its coefficients we assume the following for all $\epsilon \in (0, 1]$.

**Assumption B.2.7.** [Well posedness and existence of invariant measure] There exist a Lyapunov function $W \in C^2(\mathbb{R}^d), W \geq 1, \lim_{|x| \to \infty} W(x) = \infty$ and positive numbers $C_1, C_2$ such that for all $\alpha > 0$ and for all $\mu \in \mathcal{M}_\alpha := \{ \mu \in \mathcal{P}(\mathbb{R}^d) | \int_{\mathbb{R}^d} W(x) d\mu(x) \leq \alpha \}$, we have $L_\mu W \leq C_1 - C_2 W,$

where $L_\mu$ is the infinitesimal operator $L$ evaluated at the measure $\mu$.

**Assumption B.2.8.** [Regularity and growth coefficients]

1. Fixed $y \in \mathbb{R}^d, \mu \in \mathcal{P}(\mathbb{R}^d)$, the coefficients $K(\cdot, y), b_0(\cdot, \mu), \sigma(\cdot)$ belong to $C^2(\mathbb{R}^d)$. Moreover, there exists a parameter $\gamma \in (0, 1/2]$ and positive functions $N_1, N_2$ such that for $b_\epsilon(x, \mu) := \int K_\epsilon(x, y)\mu(dy) + \epsilon b_0(x, \mu)$, the following is satisfied for all $\alpha > 0, \mu, \mu' \in \mathcal{M}_\alpha$:

$$|b_\epsilon(x, \mu)| \leq N_1(\alpha) W^{1/2-\gamma}(x),$$

$$|b_\epsilon(x, \mu) - b_\epsilon(x, \mu')| \leq \epsilon N_2(\alpha) W^{1/2-\gamma}(y) \int_{\mathbb{R}^d} W(z)|\mu - \mu'|(dz), \forall x \in \mathbb{R}^d.$$

For the diffusion $\sigma$ we assume it is Lipschitz and uniformly elliptic, i.e. there exist $L, \kappa > 0$ such that for all $x, y, \xi \in \mathbb{R}^d$

$$\xi \sigma^*(x)\xi^* \geq \kappa |\xi|^2, \quad |\sigma^*(x) - \sigma^*(y)| \leq L|x - y|.$$

2. Additionally, suppose that the coefficients are Hölder continuous in the space component and their derivatives satisfy the following set of growth conditions stated in terms of $b_\epsilon(s, y; x) := \int K_\epsilon(y, z)\mathcal{L}(X^\epsilon_s)(dz) + \epsilon b_0(y, \mathcal{L}(X^\epsilon_s))$ and $\sigma(x)$: there exist $C, \epsilon'>0$, such that for all $(t, y) \in [0, 1] \times \mathbb{R}^d$ and all $x \in \mathbb{R}^d$,

$$|\beta_\epsilon(t, y; x)| \leq C(1 + |y|), \quad |\partial_x \beta_\epsilon(t, y; x)| \leq C(1 + |y|)^{1-\epsilon'}, \quad |\partial^2_x \beta_\epsilon(t, y; x)| \leq C(1 + |y|)^{3-\epsilon'};$$

and

$$|\sigma(x)| \leq C, \quad |\partial_x \sigma(x)| \leq C(1 + |y|), \quad |\partial^2_x \sigma(x)| \leq C(1 + |y|)^{2-\epsilon'}.$$

Notice that depending on the shape of the Lyapunov function, the linear growth condition on $\beta_\epsilon$ might be redundant.

Finally, consider test functions within the family $\mathcal{F} = \{ \phi : \mathbb{R}^d \to \mathbb{R} | \phi(x) \leq W^T(x), \forall x \in \mathbb{R}^d \}$.

**Theorem B.2.9.** Consider $V(s, y; x) = \mathbb{E}[\phi(Y^0_{s, y, x})], \text{ where } (Y^0_{s, y, x}) \text{ is defined from } (X^\epsilon_s)_{s \geq 0} \text{ (defined in turn by (B.2.4)) as in (4.3.1) and } \phi \in \mathcal{F}. \text{ Suppose that Assumptions B.2.8 and B.2.7 are satisfied. Then, there exists } \epsilon_0 \text{ such that for each } \epsilon \in [0, \epsilon_0] \text{ and for } n = 1, 2, \text{ there exists } C, c, \nu > 0 \text{ such that}$

$$|\partial^n_x V(s, y; x)|_{|x| = \nu} \leq Ce^{-cs}(1 + e^{|x|^2}), \quad \text{ for all } (s, x) \in [1, \infty) \times \mathbb{R}^d.$$

**Proof.** For any $x \in \mathbb{R}^d, s \geq 0$, let us denote by $(s, x''') \mapsto P^1_{1+s}(s, x''')$ the density of $X^1_{1+s}$. From Assumption B.2.7 and B.2.8 we know by [30, Theorem 3.3] that there exist $\epsilon_0$ such that for each $\epsilon \in [0, \epsilon_0)$ and $y \in \mathbb{R}^d$, there exists a stationary distribution $q$ such that

$$\int_{\mathbb{R}^d} W^T(x')|P^1_{1+s}(y, dx'; x)|_{y = \epsilon} q(dx') = \int_{\mathbb{R}^d} W^T(x')|P^1_{1+s}(x, dx') - q(dx')| \leq Ce^{-cs},$$

137
for some positive, non dependent on $x$, numbers $C, c$. This together with the definition of $f$ implies that Assumptions 2.2.2 in Chapter 2 holds with $s \mapsto G(s) = Ce^{-cs}$ and constant $g$.

Notice moreover that [26, Theorem VI.5] (whose conditions are included in Assumption B.2.8) implies that Assumptions 2.2.3 in Chapter 2 holds: 

$$\int_{\mathbb{R}^d} |\partial^n_x p_1(y, x'; x) |_{y=x} \leq Ce^{y|x|^2}.$$  

Therefore, by Theorem 2.2.4, the first and second derivatives in $y$ decay exponentially.
Appendix C

Appendix Chapter 4

C.1 Strong propagation of chaos for the subsampled system

Theorem C.1.1. Given a square integrable random variable $\xi$ and an $\mathbb{R}^d$–valued function $\alpha \in C^1_b(\mathbb{R}^d)$, consider the following SDE for $t \geq 0$:

$$dX_t = \mathbb{E}[\alpha(X_t)] dt + dB_t, \quad X_0 = \xi; \quad (C.1.1)$$

and, for some $0 < S \leq N \in \mathbb{N}$, its approximating subsampled particle system

$$dZ_{i,S}^{i,N} = \frac{1}{S} \sum_{j=1}^{S} \alpha(Z_{s}^{j(S)(j),N}) dt + B_i, \quad Z_{0,S}^{i,N} = \xi_i, \quad i = 1, \ldots, N. \quad (C.1.2)$$

For $(X^i)_{i=1,\ldots,N}$, i.i.d copies of $X$, the following holds:

$$\forall i \geq 1, \forall T > 0, \sup_{N \in \mathbb{N}} \left( \mathbb{E} \left[ \sup_{t \leq T} |Z_{t,S}^{i,N} - X^i_t| \right] \right) < \infty.$$

Proof. Fix a arbitrary $i, j' \in [1, \ldots, d]$. As a consequence of $(X^i)_{i=1,\ldots,N}$ being i.i.d, the particular shape of the drift and the diffusion being constant, we have

$$\frac{1}{N} \sum_{j=1}^{N} (Z_t^{i,N,j} - X_t^{i,j'}) = \int_0^t \left( \frac{1}{S} \sum_{j=1}^{S} \alpha(Z_s^{j(S)(j),N,S}) - \int_{\mathbb{R}^d} \alpha(y) \mathcal{L}(X_t^{i,j'}) (dy) \right) ds.$$

If we follow blindly the proof by Sznitman in [55, Theorem 1.4], we will obtain a weak error of order $O(1/S^{1/2})$. This is much worse (if $S \ll N$) than the order $O(1/N^{1/2})$ obtained when approximated with the full/non–subsampled particle system. One way to go around this is to use the tower property trick (as in identity (4.2.4)) in order to recover the whole empirical measure of the subsampled system. Namely, the following holds due to linearity:

$$\mathbb{E} \left[ \sup_{0 \leq t \leq T} \left| \frac{1}{N} \sum_{j=1}^{N} (Z_t^{i,N,j} - X_t^{i,j'}) \right| \right] 
\leq \mathbb{E} \left[ \int_0^T \left| \frac{1}{S} \sum_{j=1}^{S} \alpha(Z_s^{j(S)(j),N,S}) - \int_{\mathbb{R}^d} \alpha(y) \mathcal{L}(X_t^{i,j'}) (dy) \right| ds \right]$$
Finally, by the centring of $b_s$ we therefore land on the perfect shape for applying Grönwall’s lemma in order to obtain

$$\mathbb{E}\left[ \sup_{0 \leq t \leq T} \left| \frac{1}{N} \sum_{j=1}^{N} (Z_s^j, N, S) - X^f_s \right| \right]$$

Now we are in the right shape to proceed with Sznitman’s argument and first we use the Lipschitz condition on the drift:

$$\mathbb{E}\left[ \sup_{0 \leq t \leq T} \left| \frac{1}{N} \sum_{j=1}^{N} (Z_s^j, N, S) - X^f_s \right| \right]$$

Using now symmetry twice, one obtains the following bound for the auxiliary centred function $b_s(x) := \alpha(x) - \int_{\mathbb{R}^d} \alpha(y) \mathcal{L}^f(x^f_s)(dy)$:

$$\mathbb{E}\left[ \sup_{0 \leq t \leq T} \left| Z_t^{i, N, S} - X^f_t \right| \right] = \mathbb{E}\left[ \sup_{0 \leq t \leq T} \left| \frac{1}{N} \sum_{j=1}^{N} (Z_t^{i, N, S} - X^f_t) \right| \right]$$

Finally, by the centring of $b_s$, Cauchy–Schwarz inequality and the fact that $\alpha$ (and therefore...
\(b^2_t\) is bounded,
\[
\mathbb{E}\left[\frac{1}{N} \sum_{j=1}^{N} b_s(X^j_s)\right] \leq \left(\mathbb{E}\left[\frac{1}{N} \sum_{j=1}^{N} (b_s(X^j_s))^2\right]\right)^{1/2} = \left(\mathbb{E}\left[\frac{1}{N^2} \sum_{j=1}^{N} (b_s(X^j_s))^2\right]\right)^{1/2} \leq \frac{C}{\sqrt{N}}
\]

We conclude then by putting it all together:
\[
\mathbb{E}\left[\sup_{0 \leq t \leq T} |X^i_t, N, s - X^j_t|\right] \leq e^{CT} \int_0^T \frac{C}{\sqrt{N}} ds \leq \frac{C(T)}{\sqrt{N}},
\]
for some time dependent constant \(C\), but which does not depend on \(N\) or \(S\).

d\(\square\)

We make a final remark by attracting the attention to the time dependent constant. This bound can however be uniform if the drift was time dependent (although not increasing in this component) and its Lipschitz constant was integrable over \([0, \infty)\).

### C.2 Moment bounds

**Lemma C.2.1.** Suppose that Assumptions 4.1.1 and 4.1.3 hold. Given \(\xi \in L^m(\mathbb{R}^d)\) for some \(2 \leq m \leq p^*\), consider the processes defined by (4.1.1) and (4.1.5). Then, under Assumptions 4.1.1 and 4.1.3, \(X^\xi_s = Y^{0,0,\xi}_s\) in \(L^m(\mathbb{R}^d)\) for all \(s \in [0, \infty)\).

Notice that, in particular, Lemma C.2.1 implies that \(\mathcal{V}(s, x) = \mathbb{E}[\phi(Y^{0,0,\xi}_s)] = \mathbb{E}[\phi(X^{0,\xi}_s)] = \mathcal{V}(s, x)\) for all \((s, x) \in [0, \infty) \times \mathbb{R}^d\) and all \(\phi \in B_{p^*}(\mathbb{R}^d)\) (growing at most as a polynomial of order \(p^*\)).

**Proof Lemma C.2.1.** Let \(\lambda\) be an arbitrary positive constant. If Itô’s formula is applied to the stochastic process \((e^{\lambda s}X^\xi_s - Y^{0,0,\xi}_s)^m\), where \(0 < m \leq p^*\), we obtain:
\[
d(e^{\lambda s}X^\xi_s - Y^{0,0,\xi}_s)^m
= e^{\lambda s} \left[ \lambda |X^\xi_s - Y^{0,0,\xi}_s|^m + m |X^\xi_s - Y^{0,0,\xi}_s|^{m-1} \langle X^\xi_s - Y^{0,0,\xi}_s, b(X^\xi_s, \mathcal{L}(X^\xi_s)) - b(Y^{0,0,\xi}_s, \mathcal{L}(X^\xi_s), v) \rangle + \frac{m(m-1)}{2} |X^\xi_s - Y^{0,0,\xi}_s|^{m-2} \text{tr}\left(\langle \sigma(X^\xi_s, v) - \sigma(Y^{0,0,\xi}_s, v), \sigma(X^\xi_s, v) - \sigma(Y^{0,0,\xi}_s, v)\rangle^\ast\right)\right]\]
\[
+ e^{\lambda s} \left[ m |X^\xi_s - Y^{0,0,\xi}_s|^{m-1} \langle X^\xi_s - Y^{0,0,\xi}_s, \mathcal{L}(X^\xi_s), v \rangle - b(Y^{0,0,\xi}_s, \mathcal{L}(X^\xi_s), v) \rangle + \frac{m(m-1)}{2} \mathbb{E} \left[ |X^\xi_s - Y^{0,0,\xi}_s|^{m-2} \text{tr}\left(\langle \sigma(X^\xi_s, v) - \sigma(Y^{0,0,\xi}_s, v), \sigma(X^\xi_s, v) - \sigma(Y^{0,0,\xi}_s, v)\rangle^\ast\right)\right]\right] ds
\]

And now, by taking expectations of the above a using a stopping time argument similar to that in Lemma 4.3.6, we arrive to
\[
d\mathbb{E}[e^{\lambda s}X^\xi_s - Y^{0,0,\xi}_s]^m
= e^{\lambda s} \left[ \lambda |X^\xi_s - Y^{0,0,\xi}_s|^m \right]
+ m \mathbb{E} \left[ |X^\xi_s - Y^{0,0,\xi}_s|^{m-1} \langle X^\xi_s - Y^{0,0,\xi}_s, b(X^\xi_s, \mathcal{L}(X^\xi_s), v) - b(Y^{0,0,\xi}_s, \mathcal{L}(X^\xi_s), v) \rangle \right]
+ \frac{m(m-1)}{2} \mathbb{E} \left[ |X^\xi_s - Y^{0,0,\xi}_s|^{m-2} \text{tr}\left(\langle \sigma(X^\xi_s, v) - \sigma(Y^{0,0,\xi}_s, v), \sigma(X^\xi_s, v) - \sigma(Y^{0,0,\xi}_s, v)\rangle^\ast\right)\right] ds
\]
which can be treated with the monotonicity Assumption 4.1.3 in order to obtain
\[ e^{\lambda s}(\lambda - M_1 m)\mathbb{E}[X_s^\xi - Y_s^0,\xi^m] + mM_0\mathbb{E}[X_s^\xi - X_s^\xi]^2]ds. \] (C.2.1)

Now, for the particular choice of \( \lambda = mM_1 > 0 \), (C.2.1) implies:
\[ d\mathbb{E}[e^{mM_1s}|X_s^\xi - Y_s^0,\xi^m]| \leq 0 ds, \]
which by integrating in between 0 and s becomes once more
\[ \mathbb{E}[|X_s^\xi - Y_s^0,\xi^m|] \leq e^{-mM_1s}\mathbb{E}[|\xi - \xi^m|] = 0. \]

Finally, since this holds for all s, we conclude that the two processes coincide almost surely in \( L^m \).

\( \square \)

**Lemma C.2.2.** Given \( \xi, \xi' \in L^2(\mathbb{R}^d) \) and \( \nu \in \mathcal{P}(\mathbb{R}^d) \), consider the processes defined by (4.1.1). Then, under Assumption 4.1.3, for all \( s \in [0, \infty) \) we have
\[ \mathbb{E}[|X_s^\xi - X_s^{\xi'}|^2] \leq e^{-2(M_1-M_0)s}\mathbb{E}[|\xi - \xi'|^2]. \]

**Proof.** Let \( \lambda > 0 \) be an arbitrary constant and apply Itô’s formula to the process \((e^{\lambda s}|X_s^\xi - X_s^{\xi'}|^2)_{s \geq 0}\):
\[
\begin{align*}
    d(e^{\lambda s}|X_s^\xi - X_s^{\xi'}|^2) &= e^{\lambda s}(\lambda |X_s^\xi - X_s^{\xi'}|^2 + 2\langle X_s^\xi - X_s^{\xi'}, b(X_s^\xi, \mathcal{L}(X_s^\xi), \nu) - b(X_s^{\xi'}, \mathcal{L}(X_s^{\xi'}), \nu) \rangle \\
    &+ \text{tr}(\sigma(X_s^\xi, \nu) - \sigma(X_s^{\xi'}, \nu))(\sigma(X_s^\xi, \nu) - \sigma(X_s^{\xi'}, \nu)^*)ds + e^{\lambda s}(X_s^\xi - X_s^{\xi'}, (\sigma(X_s^\xi, \nu) - \sigma(X_s^{\xi'}, \nu))dB_s).
\end{align*}
\]
By taking expectations of the above and using a similar stopping time argument to that in Lemma 4.3.6, we conclude
\[
\begin{align*}
    d\mathbb{E}[e^{\lambda s}|X_s^\xi - X_s^{\xi'}|^2] &= e^{\lambda s}(\lambda \mathbb{E}[|X_s^\xi - X_s^{\xi'}|^2] + 2\mathbb{E}[\langle X_s^\xi - X_s^{\xi'}, b(X_s^\xi, \mathcal{L}(X_s^\xi), \nu) - b(X_s^{\xi'}, \mathcal{L}(X_s^{\xi'}), \nu) \rangle] \\
    &+ \mathbb{E}[\text{tr}(\sigma(X_s^\xi, \nu) - \sigma(X_s^{\xi'}, \nu))(\sigma(X_s^\xi, \nu) - \sigma(X_s^{\xi'}, \nu)^*)]ds.
\end{align*}
\]
This can now be handled with the monotonicity Assumption 4.1.3 in order to obtain:
\[
\begin{align*}
    \leq e^{\lambda s}(\lambda - 2M_1)\mathbb{E}[|X_s^\xi - X_s^{\xi'}|^2] + 2M_0\mathbb{E}[|X_s^\xi - X_s^{\xi'}|^2]ds \\
    \leq e^{\lambda s}(\lambda - 2M_1 + 2M_0)\mathbb{E}[|X_s^\xi - X_s^{\xi'}|^2]ds.
\end{align*}
\]
Now, for the particular choice of \( \lambda = 2(M_1 - M_0) > 0 \), the above becomes:
\[ d\mathbb{E}[e^{2(M_1-M_0)s}|X_s^\xi - X_s^{\xi'}|^2] \leq 0 ds, \]
which by integrating in between 0 and s implies
\[ \mathbb{E}[|X_s^\xi - X_s^{\xi'}|^2] \leq e^{-2(M_1-M_0)s}\mathbb{E}[|\xi - \xi'|^2]. \]
Lemma C.2.3. Given $\xi, \xi' \in L^2(\mathbb{R}^d)$ and $\nu \in \mathcal{P}(\mathbb{R}^d)$, consider the processes defined by (4.1.4). Then, under Assumption 4.3.2, for all $s \in [0, \infty)$ we have

$$\mathbb{E}(|X^\xi_s - X^{\xi'}_s|^2) \leq e^{-2(M_1 - M_0)s} \left( \mathbb{E}(||\xi - \xi'||^2) - \frac{M}{M_1 - M_0} \right) + \frac{M}{M_1 - M_0}.$$ 

Proof. Let $\lambda > 0$ be an arbitrary constant and apply Itô's formula to the process $(e^{\lambda s}|X^\xi_s - X^{\xi'}_s|^2)_{s \geq 0}$:

$$d(e^{\lambda s}|X^\xi_s - X^{\xi'}_s|^2) = e^{\lambda s} \left( \lambda |X^\xi_s - X^{\xi'}_s|^2 + 2\langle X^\xi_s - X^{\xi'}_s, b(X^\xi_s, \mathcal{L}(X^\xi_s), \nu) - b(X^{\xi'}_s, \mathcal{L}(X^{\xi'}_s), \nu) \rangle ight) + \mathbb{E}[tr((\sigma(X^\xi_s, \mathcal{L}(X^\xi_s), \nu) - \sigma(X^{\xi'}_s, \mathcal{L}(X^{\xi'}_s), \nu))^2)] ds$$

By taking expectations of the above and using a similar stopping time argument to that in Lemma 4.3.6, we conclude

$$d\mathbb{E}[e^{\lambda s}|X^\xi_s - X^{\xi'}_s|^2] = e^{\lambda s} \left( \lambda \mathbb{E}(|X^\xi_s - X^{\xi'}_s|^2) + 2\mathbb{E}\langle X^\xi_s - X^{\xi'}_s, b(X^\xi_s, \mathcal{L}(X^\xi_s), \nu) - b(X^{\xi'}_s, \mathcal{L}(X^{\xi'}_s), \nu) \rangle ight) + \mathbb{E}[tr((\sigma(X^\xi_s, \mathcal{L}(X^\xi_s), \nu) - \sigma(X^{\xi'}_s, \mathcal{L}(X^{\xi'}_s), \nu))^2)] ds.$$

This can now be handled with the monotonicity Assumption 4.3.2 in order to obtain:

$$\leq e^{\lambda s}(\lambda \mathbb{E}(|X^\xi_s - X^{\xi'}_s|^2) + 2\mathbb{E}|X^\xi_s|^2 + 2M) ds$$

Now, for the particular choice of $\lambda = 2(M_1 - M_0) > 0$, the above becomes:

$$d\mathbb{E}[e^{2(M_1 - M_0)s}|X^\xi_s - X^{\xi'}_s|^2] \leq 2Me^{2(M_1 - M_0)s} ds,$$

which by integrating in between 0 and $s$ implies

$$\mathbb{E}(|X^\xi_s - X^{\xi'}_s|^2) \leq e^{-2(M_1 - M_0)s} \left( \mathbb{E}(||\xi - \xi'||^2) - \frac{M}{M_1 - M_0} \right) + \frac{M}{M_1 - M_0}.$$

Lemma C.2.4 (Coercivity condition). Suppose that Assumption 4.1.3 holds. Given a measure $\nu \in \mathcal{P}(\mathbb{R}^d)$, there exist $M_2 \in \mathbb{R}$ such that for any $(x, \mu) \in \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$ we have

$$\langle x, b(x, \mu, v) \rangle + \frac{(p^* - 2)}{2}|\sigma(x, \nu)|^2 \leq M_2(\nu) - \frac{M_1(\nu)}{2}|x|^2 + M_0(\nu) \int_{\mathbb{R}^d} |x|^2 d\mu(x), \quad (C.2.2)$$

where recall that $M_1 > 2M_0 > 0$.

Proof. When choosing $(x_0, \mu_0) = (0, \delta_0)$ and for any $(x, \mu) \in \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$, the monotonicity
Assumption 4.1.3 implies by the use of the Triangle and Young’s inequalities that
\[
\langle x, b(x, \mu, v) \rangle + \frac{(p^*-1)}{2} |\sigma(x, v)|^2 \leq \langle x, b(0, \delta_0, v) \rangle - M_1 |x|^2 + M_0 \int_{\mathbb{R}^d} |x|^2 d\mu(x)
\]
Hence, there exist some \( M_2(v) = \frac{1}{2M_1} |b(0, \delta_0, v)|^2 - \frac{(p^*-1)}{2} |\sigma(0, v)|^2 \in \mathbb{R} \) such that for all \( p^* \geq 2 \):
\[
\langle x, b(x, \mu, v) \rangle + \frac{p^*-2}{2} |\sigma(x, v)|^2 \leq M_2 - \frac{M_1}{2} |x|^2 + M_0 \int_{\mathbb{R}^d} |x|^2 d\mu(x).
\]
Moreover, since we can proceed with \( \sigma \) just as we did with \( b \): using generalised Young’s inequality \((ab \leq \frac{1}{\alpha^2} a^2 + \frac{\epsilon}{2} b^2 \text{ with } \epsilon = \frac{1}{p^*-1})\), we have that:
\[
\langle x, b(x, \mu, v) \rangle + \frac{p^*-1}{2} |\sigma(x, v)|^2 \leq \frac{1}{M_1} |b(0, \delta_0, v)|^2 - \frac{p^* - 1}{2} |\sigma(0, v)|^2 + \frac{p^*-1}{2} |\sigma(x, v)|^2 \leq M_1 |x|^2 + M_0 \int_{\mathbb{R}^d} |x|^2 d\mu(x).
\]

**Lemma C.2.5.** Under Assumption 4.1.3, the process \((X_s)_{s \geq 0}\) defined by (4.1.1) satisfies the following bounds for some constant \( C > 0 \) and any value of \( m \in \mathbb{N}, m \leq (p^* - 1)/2 \):
\[
E[|X_s^\xi|^m] \leq C(1 + E[|\xi|^2] e^{-(M_1 - 2M_0)s} + e^{-mM_1/2s}E[|\xi|^m]), \quad \forall s \geq 0.
\]
**Proof.** Let \( \lambda \) be an arbitrary positive constant and \( \xi \) a starting random variable. If Itô’s formula is applied to the process \((e^{\lambda s}|X_s^\xi|^m)_{s \geq 0}\), where \( m \leq (p^* - 1)/2 \), we obtain:
\[
d(e^{\lambda s}|X_s^\xi|^m) = e^{\lambda s}(\lambda|X_s^\xi|^m + m|X_s^\xi|^{m-2}\langle X_s^\xi, b(X_s^\xi, \mathcal{L}(X_s^\xi), v) \rangle
+ \frac{m(m-1)}{2} |X_s^\xi|^{m-2} \text{tr}(\sigma \sigma^*(X_s^\xi, v)) ds + e^{\lambda s}(m|X_s^\xi|^{m-2}\langle X_s^\xi, \sigma(X_s^\xi, v) dB_s \rangle).
\]
And now, by taking expectations of the above, and after using again a stopping time argument such as in Lemma 4.3.6, we arrive to
\[
dE[e^{\lambda s}|X_s^\xi|^m]
= e^{\lambda s}\left(\lambda E[|X_s^\xi|^m] + mE[|X_s^\xi|^{m-2}\langle X_s^\xi, b(X_s^\xi, \mathcal{L}(X_s^\xi), v) \rangle
+ \frac{m(m-1)}{2} E[|X_s^\xi|^{m-2} \text{tr}(\sigma \sigma^*(X_s^\xi, v)) \right] ds,
\]

144
which can be treated with the coercivity condition (C.2.2) in order to obtain
\[
\leq e^{\lambda s} \left( (\lambda - \frac{mM_1}{2})E[|X_s^\xi|^m] + mM_0 E[|X_s^\xi|^2] + mM_2 \right) ds. \tag{C.2.3}
\]

If now \( m = 2 \), we could conclude for the particular choice of \( \lambda = M_1 - 2M_0 > 0 \) that:
\[
dE\left[ e^{(M_1 - 2M_0)s} |X_s^\xi|^2 \right] \leq 2M_2 e^{(M_1 - 2M_0)s} ds,
\]
which by integrating in between 0 and \( s \) becomes
\[
E[|X_s^\xi|^2] \leq e^{-(M_1 - 2M_0)s} \left( E[|\xi|^2] - \frac{2M_2}{M_1 - 2M_0} \right) + \frac{2M_2}{M_1 - 2M_0}. \]

Now, back to an arbitrary value of \( m \geq 2 \) in (C.2.3) we have that for a different choice of parameter, namely \( \lambda = mM_1/2 > 0 \),
\[
dE\left[ e^{\frac{mM_1}{2} s} |X_s^\xi|^m \right] \leq m \left( M_2 + M_0 \frac{2M_2}{M_1 - 2M_0} \right) e^{\frac{mM_1}{2} s} + mM_0 \left( E[|\xi|^2] - \frac{2M_2}{M_1 - 2M_0} \right) e^{\left( \frac{mM_1}{2} - 1 \right) M_1 + 2M_0 s} ds,
\]
which by integrating in between 0 and \( s \) becomes
\[
E[|X_s^\xi|^m] \leq e^{-\frac{mM_1 s}{2}} E[|\xi|^m] + \left( \frac{2M_2}{M_1} + \frac{4M_0M_2}{M_1^2 - 2M_1M_0} \right) \left( 1 - e^{-\frac{mM_1 s}{2}} \right) + \left( \frac{2mM_0}{4M_0 + (m - 2)M_1} E[|\xi|^2] - \frac{4mM_0M_2}{(M_1 - 2M_0)((m - 2)M_1 + 4M_0)} \right) \left( e^{-(M_1 - 2M_0)s} - e^{-\frac{mM_1 s}{2}} \right).
\]
We conclude that for \( C = \max \left\{ 1, \frac{4M_0M_2}{M_1^2 - 2M_1M_0}, \frac{4mM_0M_2}{(M_1 - 2M_0)((m - 2)M_1 + 4M_0)} \right\} > 0 \) such that
\[
E[|X_s^\xi|^m] \leq C (1 + E[|\xi|^2]) e^{-(M_1 - 2M_0)s} + e^{-\frac{mM_1 s}{2}} E[|\xi|^m]).
\]

\[ \square \]

**Remark C.2.6.** Notice that the computation when \( M_1 \leq 0 \) can be repeated choosing \( \lambda = 0 \). In this scenario, with the help of Grönwall’s Lemma, we conclude that the moments are finite of any order less or equal to \( p^* \). However, they grow exponentially in time instead of decaying.

Next we prove the analogous result for the following Euler discretization of (4.1.1): fixed a number of steps \( M \in \mathbb{N} \) and step size \( h > 0 \) such that \( T = Mh \), recall that for
\[
\eta(t) = t_{k-1} = (k - 1)h \quad \text{if} \quad t \in [t_{k-1}, t_k), \; k = 1, \ldots, M;
\]
we consider
\[
d\tilde{X}_s = b(\tilde{X}_{\eta(s)}, \mathcal{L}(\tilde{X}_{\eta(s)})) ds + \sigma(\tilde{X}_{\eta(s)}) dB_s, \quad \tilde{X}_0 = X_0 = \xi.
\]
Notice that we omitted the dependence on the external measure $\nu$ since it plays the role of a parameter and is not affected by the discretization.

**Lemma C.2.7.** Suppose that Assumption 4.1.3 holds. Given $\xi \in L^m(\mathbb{R}^d)$, the process $(X^\xi_s)_{s \geq 0}$ defined by 4.1.4 satisfies the following bounds for some constant $C > 0$ and any $m \in \mathbb{N}$, $2 \leq m \leq (p^* - 1)/2$:

$$
\sup_{t \geq 0} \mathbb{E}[|\tilde{X}_{\eta(t)}|^m] \leq C(1 + \mathbb{E}[|\xi|^m]).
$$

**Proof.** Let $\lambda$ be an arbitrary positive constant and $X_0 = \xi \in L^m(\mathbb{R}^d)$ a starting random variable which we omit for the sake of the notation. If Itô’s formula is applied to $(e^{\lambda s}|\tilde{X}_s|^m)_{s \geq 0}$, where $2 \leq m \leq (p^* - 1)/2$, we obtain:

$$
d(e^{\lambda s}|\tilde{X}_s|^m)
= e^{\lambda s}\left(\lambda|\tilde{X}_s|^m + m|\tilde{X}_s|^{m-2}\langle \tilde{X}_s, b(\tilde{X}_{\eta(s)}, \mathcal{L}(\tilde{X}_{\eta(s)}))\rangle + \frac{m(m-1)}{2}|\tilde{X}_s|^{m-2}\text{tr}(\sigma^*(\tilde{X}_{\eta(s)}, \mathcal{L}(\tilde{X}_{\eta(s)})))\right)\]ds
+ e^{\lambda s}\left(m|\tilde{X}_s|^{m-2}\langle \tilde{X}_s, \sigma(\tilde{X}_{\eta(s)}, \mathcal{L}(\tilde{X}_{\eta(s)}))dB_s\rangle\right).
$$

And now, by taking expectations of the above, we arrive to

$$
(e^{\lambda s}\mathbb{E}[|\tilde{X}_s|^m])
= e^{\lambda s}\left(\lambda\mathbb{E}[|\tilde{X}_s|^m] + m\mathbb{E}[|\tilde{X}_s|^{m-2}\langle \tilde{X}_s, b(\tilde{X}_{\eta(s)}, \mathcal{L}(\tilde{X}_{\eta(s)}))\rangle]ight)
+ \frac{m(m-1)}{2}\mathbb{E}[|\tilde{X}_s|^{m-2}\text{tr}(\sigma^*(\tilde{X}_{\eta(s)}, \mathcal{L}(\tilde{X}_{\eta(s)})))]ds.
$$

Focusing now on $s = \eta(s)$ and appealing once more to the coercivity condition (C.2.2), we obtain

$$
d\mathbb{E}[e^{\lambda \eta(s)}|\tilde{X}_{\eta(s)}|^m] \leq e^{\lambda \eta(s)}\left(\left(\lambda - \frac{mM_1}{2}\right)\mathbb{E}[|X_{\eta(s)}|^m] + mM_0\mathbb{E}[|X_{\eta(s)}|^2] + mM_2\right)ds. \quad \text{(C.2.4)}
$$

If now $m$ was 2, we could conclude for the particular choice of $\lambda = M_1 - 2M_0 > 0$:

$$
d\mathbb{E}[e^{(M_1 - 2M_0)\eta(s)}|\tilde{X}_{\eta(s)}|^{2}] \leq 2M_2e^{(M_1 - 2M_0)\eta(s)}ds,
$$

which by integrating in between $t_{k-1}$ and $t_k$ (for any $k = 1, ..., M$) becomes:

$$
\mathbb{E}[|\tilde{X}_{t_k}|^2] \leq e^{-(M_1 - 2M_0)h}\mathbb{E}[|\tilde{X}_{t_{k-1}}|^2] + 2M_2he^{-(M_1 - 2M_0)h} \leq e^{-(M_1 - 2M_0)h}\mathbb{E}[|\tilde{X}_{t_{k-1}}|^2] + 2M_2h.
$$

If we iterate this process and recall that we can assume $0 < h < 1$:

$$
\mathbb{E}[|\tilde{X}_{t_k}|^2] \leq e^{-(M_1 - 2M_0)h}\left(e^{-(M_1 - 2M_0)h}\mathbb{E}[|\tilde{X}_{t_{k-2}}|^2] + 2M_2h\right) + 2M_2h
\leq \ldots \leq \mathbb{E}[|X_0|^2]e^{-(M_1 - 2M_0)kh} + 2M_2h \sum_{r=0}^{k-1} (e^{-(M_1 - 2M_0)h})^r
\leq \mathbb{E}[|X_0|^2]e^{-(M_1 - 2M_0)kh} + \frac{2M_2h}{1 - e^{-(M_1 - 2M_0)h}}
$$

146
\[ \leq \mathbb{E}[|X_0|^2] + \frac{2M_2}{1 - e^{-(M_1 - 2M_0)h}}. \]

When going back to an arbitrary value of \( m \geq 2 \) in (C.2.4), it becomes for the particular choice of \( \lambda = m M_1 / 2 > 0 \):

\[ d\mathbb{E}\left[e^{-\frac{m M_1}{2} \eta(s)} |\tilde{X}_{\eta(s)}|^m \right] \leq m \left( M_2 + M_0 \left( \mathbb{E}[|X_0|^2] + \frac{2M_2}{1 - e^{-(M_1 - 2M_0)h}} \right) \right) e^{-\frac{m M_1}{2} \eta(s)} ds. \] (C.2.5)

For simplicity of the notation, let us name

\[ \tilde{M}_2(h) := M_2 + M_0 \left( \mathbb{E}[|X_0|^2] + \frac{2M_2}{1 - e^{-(M_1 - 2M_0)h}} \right). \]

After integrating (C.2.5) in between \( t_{k-1} \) and \( t_k \) (again for arbitrary \( k = 1, \ldots, M \)), it can be bounded by

\[ \mathbb{E}[|\tilde{X}_{t_k}|^m] \leq e^{-\frac{mh}{2}} \mathbb{E}[|\tilde{X}_{t_{k-1}}|^m] + mh \tilde{M}_2(h). \]

Again, by iterating the computation,

\[ \mathbb{E}[|\tilde{X}_{t_k}|^m] \leq \mathbb{E}[|X_0|^m] e^{-\frac{m M_1}{2} h k} + \sum_{r=0}^{k-1} mh \tilde{M}_2(h) \left( e^{-\frac{m M_1}{2} h} \right)^r. \]

Finally, since \( h > 0 \), the above expression can be bounded as follows:

\[ \mathbb{E}[|\tilde{X}_{t_k}|^m] \leq \mathbb{E}[|X_0|^m] + \frac{mh \tilde{M}_2(h)}{1 - e^{-\frac{m M_1}{2} h}} < \infty. \]

Hence, for all \( t \in [0, T] \),

\[ \mathbb{E}[|\tilde{X}_{\eta(t)}|^m] \leq C(1 + \mathbb{E}[|\tilde{X}_0|^m]), \]

where as already insisted, the constant \( C \) is independent of \( T \). \( \square \)

The computations are the same for any of the particles in both system approximations, up to a summation over the particles trick similar to that performed in Example 4.2.5, and so we omit them on this document.

### C.3 Glossary of derivative processes

**DIMENSIONS:**

\[ \beta : \mathbb{R}^d \rightarrow \mathbb{R}^d, \quad \Sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d} \]

\[ \partial_x \beta : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}, \quad \partial_x \Sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d \times d} \]

\[ \partial^2_x \beta : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d \times d}, \quad \partial^2_x \Sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d \times d \times d} \]

**NOTATION:**

\[ A = (A^{ij})_{j=1,\ldots,d} \in \mathbb{R}^{d \times d}, \quad A^i = (A_i^{ij})_{j=1,\ldots,d} \in \mathbb{R}^{d} \]
NON-AUTONOMOUS PROCESS:

\[ dY_t^j = \beta_j(t, Y_t^x) \, dt + \sum_l \Sigma^{jl}(t, Y_t^x) \, dB_t^l; \quad \forall j = 1, \ldots, d. \]  

\[ dY_t^x = \beta(t, Y_t^x) \, dt + \Sigma(t, Y_t^x) \, dB_t, \quad Y_0^x = x \in \mathbb{R}^d. \]

**FIRST DERIVATIVE PROCESS:** It is a well defined process (see \([5, \text{Theorem 4.10}]\)).

\[ d\partial_x Y_t^j = \sum_k \partial_x Y_t^k \partial_x \beta_j(t, Y_t^x) \, dt + \sum_{k,l} \partial_x Y_t^k \partial_x \Sigma^{jl}(t, Y_t^x) \, dB_t^l; \quad \forall i, j = 1, \ldots, d. \]

\[ Z = \partial_x Y^x \in \mathbb{R}^{d \times d}, \quad Z_s^i = (Z_s^{ij})_j = (\partial_x Y_s^j) \in \mathbb{R}^d, \quad \forall i = 1, \ldots, d. \]

\[ dZ_t^{ij} = \sum_k Z_t^{ik} \partial_x \beta_j(t, Y_t) \, dt + \sum_{k,l} Z_t^{ik} \partial_x \Sigma^{jl}(t, Y_t) \, dB_t^l \]

\[ = (Z_t^i, \partial_x \beta_j(t, Y_t)) \, dt + (Z_t^i, (\partial_x \Sigma^{jl}(t, Y_t), dB_t))_k; \quad \forall i, j = 1, \ldots, d. \]  

\[ dZ_t = Z_t \partial_x \beta(t, Y_t^x) \, dt + Z_t \partial_x \Sigma(t, Y_t^x) \, dB_t, \quad Z_0 = I_d \in \mathbb{R}^{d \times d}, \]

where \(I_d \in \mathbb{R}^{d \times d}\) is the real valued, \(d\)-dimensional identity matrix.

**SECOND DERIVATIVE PROCESS:**

\[ d\partial_x \partial_x Y_t^j = \sum_{k} \partial_x \partial_x Y_t^k \partial_x \beta_j(t, Y_t^x) \, dt + \sum_{k_1, k_2} \partial_x \partial_x Y_t^{k_1} \partial_x \partial_x \Sigma^{jl}(t, Y_t^x) \, dB_t^l \]

\[ + \sum_{k, l} \partial_x \partial_x Y_t^k \partial_x \partial_x \Sigma^{jl}(t, Y_t^x) \, dB_t^l + \sum_{k_1, k_2, l} \partial_x \partial_x Y_t^{k_1} \partial_x \partial_x \partial_x \Sigma^{jl}(t, Y_t^x) \, dB_t^l \]

\[ \forall r, i, j = 1, \ldots, d. \]

\[ \tilde{Z} = \partial_x Z = \partial_x^2 Y^x \in \mathbb{R}^{d \times d \times d}, \quad \tilde{Z}_s^{ij} = (\tilde{Z}_s^{rj})_j = (\partial_x Y_s^j) \in \mathbb{R}^d. \]

\[ d\tilde{Z}_t^{rij} = \sum_{k} \tilde{Z}_t^{rjk} \partial_x \beta_j(t, Y_t) \, dt + \sum_{k_1, k_2} \tilde{Z}_t^{rjk_1} \tilde{Z}_t^{rjk_2} \partial_x \partial_x \beta_j(t, Y_t) \, dt \]

\[ + \sum_{k, l} \tilde{Z}_t^{rjk} \partial_x \Sigma^{jl}(t, Y_t) \, dB_t^l + \sum_{k_1, k_2, l} \tilde{Z}_t^{rjk_1} \tilde{Z}_t^{rjk_2} \partial_x \Sigma^{jl}(t, Y_t) \, dB_t^l \]

\[ = (\tilde{Z}_t^i, \partial_x \beta_j(t, Y_t)) + (Z_t^i, (\tilde{Z}_s^{ij}, \Sigma^{jl}(t, Y_t), dB_t))_k; \quad \forall r, i, j = 1, \ldots, d. \]

148
THIRD DERIVATIVE PROCESS:

\[ d\bar{Z} = \partial_x \bar{Z} = \partial^2_x Z = \partial^3_x Y^x \in \mathbb{R}^{d} \times \mathbb{R}^{d} \times \mathbb{R}^{d}, \quad \bar{Z}^{vri}_s = (\bar{Z}^{vri}_s)_j = (\partial_{x_i} \partial_{x_j} \partial_{x_l} Y^j)_j \in \mathbb{R}^{d} \]

\[ d\bar{Z}^{vri}_t = \]

\[ = \sum_k \bar{Z}^{vrik}_t \partial_{x_k} \beta_j(t, Y_t) dt + \sum_{k,l} \bar{Z}^{vri}_t \partial_{x_l} \partial_{x_i} \beta_j(t, Y_t) dt + \sum_{k,l} \bar{Z}^{vri}_t \partial_{x_l} \partial_{x_i} \beta_j(t, Y_t) dt \]

\[ + \sum_{k_1,k_2} \bar{Z}^{vrik}_t \partial_{x_{k_1}} \partial_{x_{k_2}} \beta_j(t, Y_t) dt + \sum_{k_1,k_2} \bar{Z}^{vri}_t \partial_{x_{k_1}} \partial_{x_{k_2}} \beta_j(t, Y_t) dt \]

\[ + \sum_{k_1,k_2,l} \bar{Z}^{vrik}_t \partial_{x_{k_1}} \partial_{x_{k_2}} \partial_{x_{l}} \beta_j(t, Y_t) dt + \sum_{k_1,k_2,l} \bar{Z}^{vri}_t \partial_{x_{k_1}} \partial_{x_{k_2}} \partial_{x_{l}} \beta_j(t, Y_t) dt \]

\[ + \sum_{k_1,k_2,l,k_3} \bar{Z}^{vrik}_t \partial_{x_{k_1}} \partial_{x_{k_2}} \partial_{x_{l}} \partial_{x_{k_3}} \beta_j(t, Y_t) dt + \sum_{k_1,k_2,l,k_3} \bar{Z}^{vri}_t \partial_{x_{k_1}} \partial_{x_{k_2}} \partial_{x_{l}} \partial_{x_{k_3}} \beta_j(t, Y_t) dt \]

\[ = \left( \langle \bar{Z}^{vri}_t, \partial_{x_j} \beta_j(t, Y_t) \rangle + \langle \bar{Z}^{vri}, (\partial_{x_j} \partial_{x_k} \beta_j(t, Y_t)) \rangle \right)_k \]

\[ + \langle \bar{Z}^{vri}, (\partial_{x_j} \partial_{x_k} \beta_j(t, Y_t)) \rangle \right)_k \]

\[ (C.34) \]

\[ + \langle \bar{Z}^{vri}, (\partial_{x_j} \partial_{x_k} \beta_j(t, Y_t)) \rangle \right)_k \]
Lemma C.3.1. Let Assumptions 4.3.1 and 4.3.2 hold. Then, $(\tilde{Z}_s)_{s \geq 0}$ defined by (C.3.4) satisfies for any $16 \leq \max\{|4m|, 8m|\} \leq (p^* - 1)/2$ the following bound in $L_{2,m,m,m}$-norm for some constants $C, c > 0$:

$$\mathbb{E}\|\tilde{Z}_s\|^m \leq Cd^{3/2}e^{-cs}, \quad \forall s \geq 0.$$ 

Proof. Before starting the proof notice that the structure of the SDE (C.3.4) is exactly the same as that of (C.3.3). Indeed, we will have a part which can be treated using the coercivity condition derived in (C.2.2) and another part which can be bounded relying on Lemmas 4.3.6, 4.3.8 and 4.3.9.

Let $\lambda$ be an arbitrary positive constant and again, let us apply Itô’s lemma to $(e^{\lambda s}|\tilde{Z}_s^{vri}|^m)_{s \geq 0}$, after which we take expectations. Using again a stopping time argument, we obtain:

$$d\mathbb{E}[e^{\lambda s}|\tilde{Z}_s^{vri}|^m] = \mathbb{E}[e^{\lambda s}\left(\lambda|\tilde{Z}_s^{vri}|^m + m|\tilde{Z}_s^{vri}|^{m-2}\sum_{j,k} \tilde{Z}_s^{vri} \tilde{Z}_s^{vrik} \partial_{x_k} \beta_j(s, Y_s)\right)$$

$$+ \frac{m(m-1)}{2}|\tilde{Z}_s^{vri}|^{m-2} \sum_{k_2, k_1, l, j} \tilde{Z}_s^{vri} \tilde{Z}_s^{vrik} \partial_{x_{k_1}} \frac{\partial}{\partial s} (s, Y_s) \tilde{Z}_s^{vrik^2} \partial_{x_{k_2}} \Sigma^{j,l}(s, Y_s)]ds$$

$$+ \mathbb{E}[e^{\lambda s}m|\tilde{Z}_s^{vri}|^{m-2} \sum_{k, k_1, j} \tilde{Z}_s^{vri} Z_s^{ik} Z_s^{vrik} \partial_{x_{k_1}} \partial_{x_k} \beta_j(s, Y_s)]$$

$$+ \mathbb{E}[e^{\lambda s}(m|\tilde{Z}_s^{vri}|^{m-2} \sum_{k_1, k_2, l, j} \tilde{Z}_s^{vri} Z_s^{ik_1} Z_s^{vrik} \partial_{x_{k_1}} \partial_{x_{k_2}} \beta_j(s, Y_s))]$$

$$+ \mathbb{E}[e^{\lambda s}(m|\tilde{Z}_s^{vri}|^{m-2} \sum_{k_1, k_2, k_3, k_4, l, j} \tilde{Z}_s^{vri} Z_s^{ik_1} Z_s^{ik_2} Z_s^{ik_3} \partial_{x_{k_1}} \partial_{x_{k_2}} \partial_{x_{k_3}} \partial_{x_{k_4}} \beta_j(s, Y_s))]$$

$$+ \mathbb{E}[e^{\lambda s}m(m-1)\frac{1}{2}|\tilde{Z}_s^{vri}|^{m-2} \sum_{j, k, l} \left[\left(\sum_{k, l} \partial_{x_i} \partial_{x_j} \partial_{x_l} Y_t^k \partial_{x_k} \Sigma^{j,l}(t, Y_t^x)dB_t^l + \sum_{k, l, k_1} \partial_{x_i} \partial_{x_j} Y_t^k \partial_{x_k} \partial_{x_{k_1}} \partial_{x_l} \Sigma^{j,l}(t, Y_t^x)dB_t^l\right)ight.$$
weak error estimates for McKean–Vlasov dynamics which depends on an external measure.

Next we apply Young’s inequality repeatedly for the remaining terms in order to isolate the \( m \)-th order moments of \((\hat{Z}_s^{\text{prior}})_j\). This, together with the bounds proved in Lemmas 4.3.6, 4.3.8 and 4.3.9 allow us to conclude that there exist \( C, c > 0 \) such that

\[
\mathbb{E}[|\hat{Z}_s^{\text{prior}}|^m] \leq Ce^{-cs}.
\]

And so, when summing over \( v, r, i \) we have the following uniform bound for some \( C, c > 0 \):

\[
\mathbb{E}[|\hat{Z}_s|^m] \lesssim C d^3 e^{-cs}.
\]

Remark C.3.2. We omit the detailed computation but the reader can observe that, through the higher order derivatives, the integrity of the structure allowing us to use the coercivity property, is preserved. And the remaining terms after applying Itô’s formula rely on the lower order derivative estimates. In particular, it holds for the forth order derivative under the condition: \( 32 \leq \max\{4ml^2, 8ml^3, 16ml^4\} \leq (p^* - 1)/2 \).

### C.4 Weak error for the Euler scheme of a McKean–Vlasov SDE

Following the proof of [89, Theorem 1.5], in Lemma C.4.1 we generalize their result to uniform weak error estimates for McKean–Vlasov dynamics which depends on an external measure.

**Lemma C.4.1.** Let Assumptions 4.1.1, 4.1.2, 4.1.3 and 4.4.1 hold for the process defined by (4.4.1). Given \( \xi \), a \( p^* \)-integrable random variable, \( T > 0 \), \( h \in (0,1) \) and \( M \in \mathbb{N} \) such that \( T = Mh \), consider its Euler scheme discretization:

\[
d\hat{X}^\xi_t = b(\hat{X}^\xi_{\eta(t)}, \mathbb{E}[\alpha(\hat{X}^\xi_{\eta(t)}), \nu]) dt + \sigma(\hat{X}^\xi_{\eta(t)}, \nu) dB_t, \quad \forall t \in [0, T]; \hat{X}^\xi_0 = \xi.
\]

Then, there exists a finite constant \( C > 0 \), independent of \( h \) or \( T \), such that

\[
\sup_{T>0} \mathbb{E}[\phi(X^\xi_T)] - \mathbb{E}[\phi(\hat{X}^\xi_T)] \leq Ch.
\]

**Proof.** Recall that we work with the specific family of drifts in the shape of \( b(x, \mu, \nu) = b(x, \mathbb{E}^\mu[\alpha(X)], \nu) \). However, in order to underline its relevance, we will only change the notation to the latter when the linear dependence in measure plays a key role.

Since \( \xi \) is fixed and \( p^* \)-integrable, let us simplify the notation as \( X = X^\xi \) and \( \hat{X} = \hat{X}^\xi \) for the remainder of the proof.

In this circumstances, the weak error resulting from considering the McKean-Vlasov’s Euler scheme (C.4.1) for an arbitrary fixed time horizon \( T > 0 \) and step size \( 0 \leq h < 1 \), can be reformulated by using Itô’s formula and PDE (4.4.4) as follows:

\[
|\mathbb{E}[\phi(\hat{X}_T)] - \mathbb{E}[\phi(X_T)]| = |\mathbb{E}[V(0, \hat{X}_T) - V(T, \hat{X}_0)]| = \left| \sum_{k=1}^{M} \mathbb{E}[V(t_k, \hat{X}_{T-t_k}) - V(t_{k-1}, \hat{X}_{T-t_{k-1}})] \right|
\]

151
\[
\sum_{k=1}^{M} \mathbb{E}\left[ \int_{t_{k-1}}^{t_k} \partial_x V(s, \tilde{X}_{T-s}) - \langle \partial_x V(s, \tilde{X}_{T-s}), b(\tilde{X}_{\eta}(s), \mathcal{L}(\tilde{X}_{\eta}(s)), v) \rangle \right. \\
\left. - \frac{1}{2} \text{tr}(\sigma(\tilde{X}_s, v) \partial_x^2 V(s, \tilde{X}_{T-s}, v) \sigma^*(\tilde{X}_s, v)) ds \right] \\
= \sum_{k=1}^{M} \mathbb{E}\left[ \int_{t_{k-1}}^{t_k} \langle \partial_x V(s, \tilde{X}_{T-s}), b(\tilde{X}_{\eta}(s), \mathcal{L}(\tilde{X}_{\eta}(s)), v) - b(\tilde{X}_s, \mathcal{L}(\tilde{X}_s), v) \rangle \right. \\
\left. + \frac{1}{2} \text{tr}(\sigma(\tilde{X}_{\eta}(s), v) \partial_x^2 V(s, \tilde{X}_{T-s}) \sigma^*(\tilde{X}_{\eta}(s), v) - \sigma(\tilde{X}_s, v) \partial_x^2 V(s, \tilde{X}_{T-s}) \sigma^*(\tilde{X}_s, v)) ds \right] \\
= \mathbb{E}\left[ \int_0^T \langle \partial_x V(s, \tilde{X}_{T-s}), b(\tilde{X}_{\eta}(s), \mathcal{L}(\tilde{X}_{\eta}(s)), v) - b(\tilde{X}_s, \mathcal{L}(\tilde{X}_s), v) \rangle ds \right] \\
+ \mathbb{E}\left[ \int_0^T \frac{1}{2} \text{tr}(\sigma(\tilde{X}_{\eta}(s), v) \partial_x^2 V(s, \tilde{X}_{T-s}) \sigma^*(\tilde{X}_{\eta}(s), v) - \sigma(\tilde{X}_s, v) \partial_x^2 V(s, \tilde{X}_{T-s}) \sigma^*(\tilde{X}_s, v)) ds \right]. \\
\text{(C.4.2)}
\]

First we notice that
\[
\mathbb{E}\left[ \int_0^T \langle \partial_x V(s, \tilde{X}_{T-s}) - \partial_x V(s, \tilde{X}_{T-\eta(s)}), b(\tilde{X}_{\eta}(s), \mathcal{L}(\tilde{X}_{\eta}(s)), v) - b(\tilde{X}_s, \mathcal{L}(\tilde{X}_s), v) \rangle ds \right] \\
\leq \mathbb{E}\left[ \int_0^T \left( \langle \partial_x V(s, \tilde{X}_{T-s}), b(\tilde{X}_{\eta}(s), \mathcal{L}(\tilde{X}_{\eta}(s)), v) - b(\tilde{X}_s, \mathcal{L}(\tilde{X}_s), v) \rangle \right. \\
\left. + \frac{1}{2} \text{tr}(\sigma(\tilde{X}_{\eta}(s), v) \partial_x^2 V(s, \tilde{X}_{T-s}) \sigma^*(\tilde{X}_{\eta}(s), v) - \sigma(\tilde{X}_s, v) \partial_x^2 V(s, \tilde{X}_{T-s}) \sigma^*(\tilde{X}_s, v)) ds \right] \right]. \\
\text{(C.4.5)}
\]

Now, recall that \( b \) satisfies the growth condition \( (4.18) \) as a consequence of Assumption \( 4.1.1 \). This, together with the decay of the derivatives of the function \( V \) from Proposition \( 4.3.10 \) (see also Remark 4.3.11), lead us to
\[
\mathbb{E}\left[ \int_0^T \langle \partial_x V(s, \tilde{X}_{T-s}) - \partial_x V(s, \tilde{X}_{T-\eta(s)}), b(\tilde{X}_{\eta}(s), \mathcal{L}(\tilde{X}_{\eta}(s)), v) - b(\tilde{X}_s, \mathcal{L}(\tilde{X}_s), v) \rangle ds \right] \\
\leq C \mathbb{E}\left[ \left( \int_0^T \left( \langle \partial_x^2 V(s, \alpha \tilde{X}_{T-s} + (1-\alpha) \tilde{X}_{T-\eta(s)}, \mathcal{L}(\tilde{X}_{\eta}(s)), v) - b(\tilde{X}_s, \mathcal{L}(\tilde{X}_s), v) \rangle \right. \\
\left. \left. \times \left( |\tilde{X}_s - \tilde{X}_{\eta(s)}|^p + |\tilde{X}_{\eta(s)}|^p \right) ds \right) \right] \right]. \\
\text{(C.4.6)}
\]
Analogously, one obtains
\[ E \left[ \int_0^T \left| \partial_s^2 V(s, \alpha \tilde{X}_{T-s} + (1 - \alpha) \tilde{X}_{T-\eta(s)} \right|^2 HS \, da \right] \leq C \int_0^T e^{-cs} \left[ (1 + |\tilde{X}_{T-s}|^{m^*/2} + |\tilde{X}_{T-\eta(s)}|^{m^*/2}) \right] \left( 1 + |\tilde{X}_s|^{p/2} + |\tilde{X}_{\eta(s)}|^{p/2} \right) \times \max(|\tilde{X}_s - \tilde{X}_{\eta(s)}|^2, |\tilde{X}_{T-s} - \tilde{X}_{T-\eta(s)}|^2) \, ds \]
\[ \leq C \int_0^T e^{-cs} \left[ (1 + |\tilde{X}_{T-s}|^{m^*/2} + |\tilde{X}_{T-\eta(s)}|^{m^*/2}) \right] \left( 1 + |\tilde{X}_s|^{p/2} + |\tilde{X}_{\eta(s)}|^{p/2} \right) \times \max(|\tilde{X}_s - \tilde{X}_{\eta(s)}|^2, |\tilde{X}_{T-s} - \tilde{X}_{T-\eta(s)}|^2) \, ds \]
\[ \leq C \int_0^T e^{-cs} \left[ (1 + |\tilde{X}_{T-s}|^{m^*/2} + |\tilde{X}_{T-\eta(s)}|^{m^*/2}) \right] \left( 1 + |\tilde{X}_s|^p + |\tilde{X}_{\eta(s)}|^p \right) \times \max\left( \mathbb{E}[|\tilde{X}_s - \tilde{X}_{\eta(s)}|^4] \right)^{1/2}, \left( \mathbb{E}[|\tilde{X}_{T-s} - \tilde{X}_{T-\eta(s)}|^4] \right)^{1/2} ds. \]

By Lemma C.2.7 (notice that it holds for moments up to order \((p^* - 1)/2 > p + m^* > 4\)), there exists \(C > 0\) independent of \(T\) such that \(\sup_{s \geq 0} \mathbb{E}[|\tilde{X}_s|] \leq C\). Therefore, (C.4.5) is bounded uniformly in \(T\). Moreover, notice that for some other constant, \(\mathbb{E}[|\tilde{X}_s - \tilde{X}_{\eta(s)}|^4] \leq C h^2\).

Indeed, the explicit form of the forth moment of a normally distributed random variable, Cauchy–Schwarz inequality applied twice, the growth of \(b\) proved in (4.1.8) and the uniform moments bounds from Lemma C.2.7, allow us to conclude:

\[ \mathbb{E}[|\tilde{X}_s - \tilde{X}_{\eta(s)}|^4] \leq \mathbb{E}\left[ \left( \int_{\eta(s)}^s b(\tilde{X}_{\eta(t)}, \mathcal{L}(\tilde{X}_{\eta(t)}), \nu) \, dt + B(s - B(\eta(s)) \right)^4 \right] \]
\[ \leq C \mathbb{E}\left[ \left( \int_{\eta(s)}^s b(\tilde{X}_{\eta(t)}, \mathcal{L}(\tilde{X}_{\eta(t)}), \nu) \, dt \right)^4 + \mathbb{E}[|B(s) - B(\eta(s))|^4] \right] \]
\[ \leq (s - \eta(s))^2 \mathbb{E}\left[ \left( \int_{\eta(s)}^s |b(\tilde{X}_{\eta(t)}, \mathcal{L}(\tilde{X}_{\eta(t)}), \nu)|^2 \, dt \right) + C 3(s - \eta(s))^2 \right] \]
\[ \leq (s - \eta(s))^3 \int_{\eta(s)}^s \mathbb{E}[|b(\tilde{X}_{\eta(t)}, \mathcal{L}(\tilde{X}_{\eta(t)}), \nu)|^4] \, dt + C 3(s - \eta(s))^2 \]
\[ \leq C((s - \eta(s))^3 \int_{\eta(s)}^s \left( 1 + \mathbb{E}[|\tilde{X}_{\eta(t)}|^{2p+4}] + \mathbb{E}[|\tilde{X}_{\eta(t)}|^8] \right) \, dt + C(s - \eta(s))^2 \]
\[ \leq C h^2. \]

Analogously, one obtains \(\mathbb{E}[|\tilde{X}_{T-s} - \tilde{X}_{T-\eta(s)}|^4] \leq C h^2\). All together we obtain the uniform estimate \(\sup_{T \geq 0} (C.4.5) \leq C h^2\).

Next, by the tower property and the compatibility theorems which under our assumptions allow us to adapt the conditional expectation inside the integral (see Section 4.2.1 for more detail or alternatively [35]),

\[ (C.4.6) = \mathbb{E}\left[ \int_0^T \left( \partial_s V(s, \tilde{X}_{T-\eta(s)}, b(\tilde{X}_{\eta(s)}, \mathcal{L}(\tilde{X}_{\eta(s)}), \nu) - b(\tilde{X}_s, \mathcal{L}(\tilde{X}_{\eta(s)}), \nu) \right) \, ds \right] \]
\[ = \mathbb{E}\left[ \int_0^T \left( \partial_s V(s, \tilde{X}_{T-\eta(s)}), b(\tilde{X}_{\eta(s)}, \mathcal{L}(\tilde{X}_{\eta(s)}), \nu) - b(\tilde{X}_s, \mathcal{L}(\tilde{X}_{\eta(s)}), \nu) \right) \, ds \right] \]
\[ \leq C \int_0^T e^{-cs} \left[ (1 + |\tilde{X}_{T-\eta(s)}|^{m^*/2}) \right] \left( b(\tilde{X}_{\eta(s)}, \mathcal{L}(\tilde{X}_{\eta(s)}), \nu) - b(\tilde{X}_s, \mathcal{L}(\tilde{X}_{\eta(s)}), \nu) \right) \, ds. \]
\[ \leq C \int_0^T e^{-cs} \left( \mathbb{E} \left[ |1 + |\tilde{X}_{T-\eta(t)}|^m| \right] \right)^{1/2} \times \left( \mathbb{E} \left[ \left| b(\tilde{X}_{\eta(t)}, \mathcal{L}(\tilde{X}_{\eta(t)}), \nu) - b(\tilde{X}_s, \mathcal{L}(\tilde{X}_{\eta(s)}), \nu) \right|^2 \right] \right)^{1/2} ds. \]

Now, by freezing the second and third components in \( b \) we can use the polynomial growth of \( \partial_x b \) and \( \partial_x^2 b \) (see Assumption 4.1.1) and the growth for \( b \) together with the linear growth of \( \sigma \), proved in (4.1.8) and (4.1.9) respectively, as follows: for any \( i \in \{1, \ldots, d\} \),

\[
b_i(\tilde{X}_{\eta(t)}, \mathcal{L}(\tilde{X}_{\eta(t)}), \nu) - b_i(\tilde{X}_s, \mathcal{L}(\tilde{X}_{\eta(s)}), \nu) = \int_{\eta(t)}^s \left\langle \partial_x b_i(\tilde{X}_t, \mathcal{L}(\tilde{X}_{\eta(t)}), \nu), b(\tilde{X}_{\eta(t)}, \mathcal{L}(\tilde{X}_{\eta(t)}), \nu) \right\rangle dt + \frac{1}{2} \text{tr} (\sigma(\tilde{X}_{\eta(t)}), \nu) \partial_x^2 b_i(\tilde{X}_t, \mathcal{L}(\tilde{X}_{\eta(t)}), \nu) \sigma^* (\tilde{X}_{\eta(t)}), \nu) dt \leq C \int_{\eta(t)}^s (1 + |\tilde{X}_t|^{p/2}) \left| b(\tilde{X}_{\eta(t)}, \mathcal{L}(\tilde{X}_{\eta(s)}), \nu) \right| ds + C \int_{\eta(t)}^s (1 + |\tilde{X}_t|^{p/2}) \left| \sigma^*(\tilde{X}_{\eta(t)}), \nu \right|_{HS} dt \leq C \int_{\eta(t)}^s (1 + |\tilde{X}_t|^{p/2})(1 + |\tilde{X}_{\eta(t)}|^{p+1} + \mathbb{E} |\tilde{X}_{\eta(t)}|^2) dt + C \int_{\eta(t)}^s (1 + |\tilde{X}_t|^{p/2}) |\tilde{X}_{\eta(t)}|^2 dt. \quad (C.4.7)\]

This means that when we take the expectation of the square in (C.4.7), due again to uniform moments proved in Lemma C.2.7 and Cauchy–Schwarz inequality, we obtain the following:

\[
\left( \mathbb{E} \left[ \left| b_i(\tilde{X}_{\eta(t)}, \mathcal{L}(\tilde{X}_{\eta(t)}), \nu) - b_i(\tilde{X}_s, \mathcal{L}(\tilde{X}_{\eta(s)}), \nu) \right|^2 \right] \right)^{1/2} \leq C \mathbb{E} \left[ \left( \int_{\eta(t)}^s (1 + |\tilde{X}_t|^{p})(1 + |\tilde{X}_{\eta(t)}|^{p+2}) dt \right)^2 \right]^{1/2} \leq C h,
\]

which together with the uniform moment bounds (Lemma C.2.7), allow us to conclude that \( \sup_{T \geq 0} (C.4.6) \leq C h \).

When bounding (C.4.3), the linearity in measure is key for the future use of Grönwall’s Lemma, so we underline it by a change in notation. For this term we apply the Lipschitz assumption on the measure component of \( b \) (see Assumption 4.1.1) and again the exponential decay in time of \( \partial_x V \) (Proposition 4.3.10) together with uniform in time moments bounds (Lemma C.2.7) in order to obtain:

\[
(C.4.3) \leq C \int_0^T e^{-cs} \mathbb{E} \left[ \left( 1 + |\tilde{X}_{T-s}|^{m/2} \right) |\mathbb{E}[\alpha(\tilde{X}_{\eta(s)})] - \mathbb{E}[\alpha(X_s)]| \right] ds \leq C \int_0^T e^{-cs} \mathbb{E}[\alpha(\tilde{X}_{\eta(s)})] - \mathbb{E}[\alpha(X_s)]| ds.
\]

As a quick summary, when gathering it all, one concludes that

\[
|\mathbb{E}[\phi(\tilde{X}_T)] - \mathbb{E}[\phi(X_T)]| \leq C \left( h + \int_0^T e^{-cs} \mathbb{E}[\alpha(\tilde{X}_{\eta(s)})] - \mathbb{E}[\alpha(X_s)]| ds \right) \quad (C.4.4).
\]

\[ 154 \]
Next notice that for the term (C.4.4) we need to bound appropriately the difference of $\sigma^*(\bar{X}_s, \nu)$ when evaluated in the process $\bar{X}_s$ and when evaluated in $\bar{X}_{\eta}(\cdot)$. First, algebraic manipulations allow us to rewrite

$$\sigma^*(\bar{X}_{\eta}(\cdot), \nu) - \sigma^*(\bar{X}_s, \nu)$$

$$= (\sigma(\bar{X}_{\eta}(\cdot), \nu) - \sigma(\bar{X}_s, \nu))(\sigma^*(\bar{X}_{\eta}(\cdot), \nu) - \sigma^*(\bar{X}_s, \nu))$$

$$+ \sigma(\bar{X}_{\eta}(\cdot), \nu)^\sigma*(\bar{X}_s, \nu) + \sigma(\bar{X}_s, \nu)\sigma^*(\bar{X}_{\eta}(\cdot), \nu) - 2\sigma(\bar{X}_s, \nu)\sigma^*(\bar{X}_s, \nu)$$

$$= (\sigma(\bar{X}_{\eta}(\cdot), \nu) - \sigma(\bar{X}_s, \nu))(\sigma^*(\bar{X}_{\eta}(\cdot), \nu) - \sigma^*(\bar{X}_s, \nu))$$

$$+ \sigma(\bar{X}_s, \nu)(\sigma^*(\bar{X}_{\eta}(\cdot), \nu) - \sigma^*(\bar{X}_s, \nu)) + (\sigma(\bar{X}_{\eta}(\cdot), \nu) - \sigma(\bar{X}_s, \nu))\sigma^*(\bar{X}_s, \nu).$$

Based on the above decomposition and given that $\sigma$ is actually Lipschitz because its derivative is bounded (Assumption 4.1.1) and it moreover satisfies the growth condition (4.1.9), we obtain the following bounds on the variance of $\sigma^*(\bar{X}_{\eta}(\cdot), \nu)$ as a consequence of the triangle inequality, Cauchy–Schwarz inequality and the moments bounds presented in Lemma C.2.7:

$$\mathbb{E}\left[||\sigma^*(\bar{X}_{\eta}(\cdot), \nu) - \sigma^*(\bar{X}_s, \nu)||^2_{HS}\right]$$

$$\leq \mathbb{E}\left[||\sigma(\bar{X}_{\eta}(\cdot), \nu) - \sigma(\bar{X}_s, \nu)||^4_{HS}\right]$$

$$+ \left(\mathbb{E}\left[||\sigma(\bar{X}_{\eta}(\cdot), \nu) - \sigma(\bar{X}_s, \nu)||^4_{HS}\right]\right)^{1/2}\left(\mathbb{E}\left[||\sigma(\bar{X}_s, \nu)||^4_{HS}\right] + \mathbb{E}\left[||\sigma(\bar{X}_{\eta}(\cdot), \nu)||^4_{HS}\right]\right)^{1/2}$$

(C.4.9)

$$\leq C\left(\mathbb{E}\left[\bar{X}_{\eta}(\cdot) - \bar{X}_s|^4\right] + \left(\mathbb{E}\left[1 + |\bar{X}_{\eta}(\cdot)|^4\right]\right)^{1/2}\left(\mathbb{E}\left[|\bar{X}_{\eta}(\cdot) - \bar{X}_s|^4\right]\right)^{1/2}\right)$$

$$\leq C h^2.$$

After using Cauchy–Schwarz inequality, the bound in (C.4.9) and and again the bounds on the moments, one obtains:

(C.4.4) $\leq \mathbb{E}\left[\int_0^T ||\partial_s^V(s, \bar{X}_{T-s})||_{HS}\times||\sigma^*(\bar{X}_{\eta}(\cdot), \nu) - \sigma^*(\bar{X}_s, \nu)||_{HS} ds\right]$$

$$\leq C \int_0^T \mathbb{E}\left[||\partial_s^V(s, \bar{X}_{T-s})||^2_{HS}\right]^{1/2}\left(\mathbb{E}\left[||\sigma^*(\bar{X}_{\eta}(\cdot), \nu) - \sigma^*(\bar{X}_s, \nu)||^2_{HS}\right]\right)^{1/2} ds$$

$$\leq C h \int_0^T e^{-cs}\left(\mathbb{E}\left[1 + |\bar{X}_{T-s}|^{m}\right]\right)^{1/2} ds$$

$$\leq C h.$$

Taking the bound of the term (C.4.4) back into (C.4.8), we conclude that

$$|\mathbb{E}[\phi(\bar{X}_{T})] - \mathbb{E}[\phi(X_T)]| \leq C\left(h + \int_0^T e^{-cs}|E[\alpha(\bar{X}_{\eta}(\cdot))]| - |E[\alpha(X_3)]| ds\right).$$

(C.10.4)

Notice that for the particular test function $\phi = \alpha$, (C.10.4) and Grönwall’s inequality imply that

$$|\mathbb{E}[\alpha(\bar{X}_{\eta}(\cdot))] - \mathbb{E}[\alpha(X_3)]| \leq C h e^{\int_0^T e^{-cs} ds} \leq C h,$$

where the constant $C$ is independent of $T$. Finally, the uniformly in time bound of order $\mathcal{O}(h)$ for the weak error follows from taking the previous bound back into (C.10.4). □
C.5 Weak error induced by the discretization of the full particle system

This section studies, inspired by [51], the error induced by considering an Euler scheme of a particle system (C.5.1), in contraposition with Lemma C.4.1, which considers the discretization of the McKean–Vlasov SDE (4.4.1).

Proposition C.5.1. Fix $N \in \mathbb{N}$ and let Assumptions 4.1.1, 4.1.2, 4.1.3 and 4.4.1 hold. Given $\xi$, a $p^*$–integrable random variable, let $(X_t^i)_{t \geq 0}$ satisfy (4.4.1). When approximated by the following particle system:

\[
\left\{
\begin{array}{ll}
    dY_t^{i,N} = b(Y_t^{i,N}, \mu_t^{Y,N}, v)dt + \sigma(Y_t^{i,N}, v)dB_t^i, & \forall t \in [0, \infty); i = 1, ..., N; \\
    Y_0^{i,N} = \xi^i,
\end{array}
\right.
\] (C.5.1)

whose Euler scheme is given, up to some time $0 < T = Mh$ and for chosen $h > 0, M \in \mathbb{N}$, by

\[
\left\{
\begin{array}{ll}
    d\tilde{Y}_t^{i,N} = b(\tilde{Y}_t^{i,N}, \mu_{\eta(t)}^{Y,N}, v)dt + \sigma(\tilde{Y}_t^{i,N}, v)dB_t^i, & \forall t \in [0, T]; i = 1, ..., N; \\
    \tilde{Y}_0^{i,N} = \xi^i,
\end{array}
\right.
\] (C.5.2)

there exists a finite constant $C > 0$, independent of $h, N$ or $T$, such that

\[
\sup_{T \geq 0} \left| \mathbb{E}[\phi(X_T^\xi)] - \mathbb{E}[\phi(\tilde{Y}_T^{1,N})] \right| \leq C(h + 1/N).
\]

Proof. Again, since $\xi$ is fixed and $p^*$–integrable, let us simplify the notation as $X = X^\xi$ for the remainder of the proof.

Just as in the previous sections, the weak error can be recast as $\left| \mathbb{E}[\phi(X_T)] - \mathbb{E}[\phi(\tilde{Y}_T^{1,N})] \right|$. Moreover, recall that we defined for values $(t, x) \in [0, \infty) \times \mathbb{R}^d$ the functions $V(t, x) = \mathbb{E}[\phi(X_t^x)]$ and the associated to it through (4.3.1), $V(t, x; \xi) = \mathbb{E}[\phi(Y_t^{x,\xi})]$. They are proved in Lemma C.2.1 to satisfy $V(t, x) = V(t, x; x)$ and, as a consequence, bounds of the weak error can be attained by using Itô’s formula and the PDE (4.1.6) as follows:

\[
\left| \mathbb{E}[\phi(\tilde{Y}_T^{1,N})] - \mathbb{E}[\phi(X_T)] \right| = \left| \mathbb{E}[V(T, \tilde{Y}_0^{1,N}) - V(0, \tilde{Y}_T^{1,N})] \right|
\]

\[
= \left| \sum_{k=1}^M \mathbb{E}\left[ V(t_k, \tilde{Y}_{T-k}^{1,N}) - V(t_{k-1}, \tilde{Y}_{T-k-1}^{1,N}) \right] \right|
\]

\[
= \left| \sum_{k=1}^M \mathbb{E}\left[ \int_{t_{k-1}}^{t_k} \partial_s V(s, \tilde{Y}_{T-s}^{1,N}) + \langle D V(s, \tilde{Y}_{T-s}^{1,N}), b(\tilde{Y}_{\eta(s)}^{1,N}, \mu_{\eta(s)}^{Y,N}, v) \rangle \right. \right.
\]

\[
\left. + \frac{1}{2} \text{tr}\left( \sigma(\tilde{Y}_{\eta(s)}^{1,N}, v)\partial^2_s V(s, \tilde{Y}_{T-s}^{1,N})\sigma^*(\tilde{Y}_{\eta(s)}^{1,N}, v) \right) ds \right] \right|
\]

\[
= \left| \mathbb{E}\left[ \int_0^T \partial_s V(s, \tilde{Y}_{T-s}^{1,N}) + \langle \partial_s V(s, \tilde{Y}_{T-s}^{1,N}), b(\tilde{Y}_{\eta(s)}^{1,N}, \mu_{\eta(s)}^{Y,N}, v) \rangle \right. \right.
\]

\[
\left. + \frac{1}{2} \text{tr}\left( \sigma(\tilde{Y}_{\eta(s)}^{1,N}, v)\partial^2_s V(s, \tilde{Y}_{T-s}^{1,N})\sigma^*(\tilde{Y}_{\eta(s)}^{1,N}, v) \right) ds \right] \right|
\]

\[
= \left| \mathbb{E}\left[ \int_0^T \langle \partial_x V(s, \tilde{Y}_{T-s}^{1,N}), b(\tilde{Y}_{\eta(s)}^{1,N}, \mu_{\eta(s)}^{Y,N}, v) - b(\tilde{Y}_s^{1,N}, \mathcal{L}(X_s, v)) \rangle ds \right] \right|
\]

156
4.1.3 and here it proves useful again in order to understand the implications of our assumption 4.1.1 this means that measure–space derivatives and Jensen’s inequality can be used to

Moreover, for the particular measure dependence of the drift considered in (4.4.1) we know

P

prehensible introduction to calculus on

measure derivative

δ

x

tions. Indeed, due to Assumption 4.1.1, for any (b

µ

We already mentioned the notion of derivative with respect to the measure in Example 4.1.3 and here it proves useful again in order to understand the implications of our assumptions. Indeed, due to Assumption 4.1.1, for any \((x, \mu, \nu) \in \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d)\) the linear measure derivative \(\frac{\partial b}{\partial \mu}(x, \mu, \nu, \cdot)\) and the intrinsic derivative \(D_{\mu} b(x, \mu, \nu, \cdot)\) exist (for a comprehensible introduction to calculus on \(\mathcal{P}(\mathbb{R}^d)\) see [54, Section 9] and [35, Ch. 5, Vol. I]). Moreover, for the particular measure dependence of the drift considered in (4.4.1) we know that they are equal to \(\partial_{x} b(x, \mathbb{E}[\alpha(X)], \nu)\alpha(\cdot)\) and \(\partial_{x} b(x, \mathbb{E}[\alpha(X)], \nu)\alpha(\cdot)\) respectively, where \(X \sim \mu\) (see [41]) and they are bounded under Assumption 4.4.1. Together with Assumption 4.1.1 this means that measure–space derivatives and Jensen’s inequality can be used to conclude:

\[
\begin{align*}
&\mathbb{E}\left[ b(\bar{Y}_{\nu}^{1,N} \nu), \mathcal{L}(X_{\nu})-b(\bar{Y}_{\nu}^{1,N} \nu), \mathcal{L}(X_{\nu}) \right] \\
&= \mathbb{E}\left[ b(\bar{Y}_{\nu}^{1,N} \nu, \mathbb{E}[\mathcal{L}(X_{\nu})], \nu)-b(\bar{Y}_{\nu}^{1,N} \nu, \mathbb{E}[\mathcal{L}(X_{\nu})], \nu) \right] \\
&\leq \mathbb{E}\left[ \left| \int_{0}^{1} \mathbb{E} \left[ \mathcal{L}(X_{\nu}) \right] - b \left( \bar{Y}_{\nu}^{1,N} \nu, \mathcal{L}(X_{\nu}) \right) (d\nu)dt \right| \right] \end{align*}
\]
we conclude:

4.3.10 and Remark 4.3.11) and finally the bound in (C.5.9) and the uniform moments estimates

Using again Cauchy–Schwarz inequality, exponential decay in time for

A

\int_{\eta(s)}^{T} \left| \bar{Y}_{s}^{1,N} \right| d s

= \left| E \left[ \int_{\eta(s)}^{T} \partial_x V(s, \bar{Y}_{s}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) d s \right] \right|

≤ C e^{-c s} \mathbb{E} \left[ \left( 1 + \left| \bar{Y}_{s}^{1,N} \right|^{p} + \left| \bar{Y}_{s}^{1,N} \right|^{p} \right) \right] \left( \left| \bar{Y}_{s}^{1,N} \right| - \bar{Y}_{\eta(s)}^{1,N} \right) \left( \left| \bar{Y}_{s}^{1,N} \right| \right) d s

≤ e^{-c s} \mathbb{E} [ |\alpha(X_s) - \alpha(X_{\eta(s)})|^2] d s.

Now we are ready to consider each of the terms from A_1 to A_5. Using first Cauchy–Schwarz inequality, afterwards that \partial_x V is bounded with exponential decay in time (see Proposition 4.3.10 and Remark 4.3.11) and finally the bound in (C.5.9) and the uniform moments estimates from Lemma C.2.7,

A_1 = \left| E \left[ \int_{0}^{T} \langle \partial_x V(s, \bar{Y}_{s}^{1,N}), b(\bar{Y}_{s}^{1,N}, \mathcal{L}(X_{s}), \nu) - b(\bar{Y}_{s}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) \rangle d s \right] \right|

≤ \int_{0}^{T} \mathbb{E} \left[ \left| \partial_x V(s, \bar{Y}_{s}^{1,N}), b(\bar{Y}_{s}^{1,N}, \mathcal{L}(X_{s}), \nu) - b(\bar{Y}_{s}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) \right| d s \right]

≤ C \int_{0}^{T} e^{-c s} \mathbb{E} \left[ \left( 1 + \left| \bar{Y}_{s}^{1,N} \right|^{m/2} \right) \left| b(\bar{Y}_{s}^{1,N}, \mathcal{L}(X_{s}), \nu) - b(\bar{Y}_{\eta(s)}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) \right| d s \right]

≤ C \int_{0}^{T} e^{-c s} \mathbb{E} \left[ \left( 1 + \left| \bar{Y}_{s}^{1,N} \right|^{m} \right) \right] \left( \left| b(\bar{Y}_{s}^{1,N}, \mathcal{L}(X_{s}), \nu) - b(\bar{Y}_{\eta(s)}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) \right| \right)^{1/2} d s

≤ C \int_{0}^{T} e^{-c s} \mathbb{E} \left[ |\alpha(X_s) - \alpha(X_{\eta(s)})|^2 \right] d s.

Using again Cauchy–Schwarz inequality, exponential decay in time for \partial_x V, the Lipschitz condition on the drift in Assumption 4.1.1 and finally that the moments are uniformly bounded, we conclude:

A_2 = \left| E \left[ \int_{0}^{T} \langle \partial_x V(s, \bar{Y}_{s}^{1,N}), - \partial_x V(s, \bar{Y}_{s}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) - b(\bar{Y}_{\eta(s)}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) \rangle d s \right] \right|

≤ C \int_{0}^{T} e^{-c s} h^{1/2} \left( \mathbb{E} \left[ \left( 1 + \left| \bar{Y}_{s}^{1,N} \right|^{p} + \left| \bar{Y}_{s}^{1,N} \right|^{p} \right) \right] \left( \left| \bar{Y}_{s}^{1,N} \right| - \bar{Y}_{\eta(s)}^{1,N} \right) \right)^{1/2} d s

≤ C h.

Using a second order Taylor expansion for b(\bar{Y}_{s}^{1,N}, \cdot, \cdot), the polynomial growth of b, \partial_x b and \partial^2_x b (see Assumption 4.1.1, display (4.1.8)) and the linear growth of \sigma, the fact that \partial_x V has exponential decay in time and we have uniform moments bounds (see Appendix C.2), we conclude finally with Hölder inequality’s help:

A_3 = \left| E \left[ \int_{0}^{T} \langle \partial_x V(s, \bar{Y}_{s}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) - b(\bar{Y}_{\eta(s)}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) \rangle d s \right] \right|

= \left| E \left[ \int_{0}^{T} \langle \partial_x V(s, \bar{Y}_{s}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) \right] \left( \int_{\eta(s)}^{s} \langle \partial_x b(\bar{Y}_{t}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu), b(\bar{Y}_{t}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) \rangle + \frac{1}{2} \text{tr}(\sigma(\bar{Y}_{t}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) \sigma(\bar{Y}_{t}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) \mathcal{L}(X_{\eta(s)}), \nu) \sigma(\bar{Y}_{t}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) \rangle d t \right) d s \right] \right|

= \left| E \left[ \int_{0}^{T} \langle \partial_x V(s, \bar{Y}_{s}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) \right] \left( \int_{\eta(s)}^{s} \langle \partial_x b(\bar{Y}_{t}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu), b(\bar{Y}_{t}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) \rangle \right) d s \right] \right|
\[
+ \frac{1}{2} \text{tr}\left( \sigma(\tilde{Y}_{t}^{1,N}, \nu) \partial_x \partial_x b(\tilde{Y}_{t}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) \sigma^*(\tilde{Y}_{t}^{1,N}, \nu) \right) dt \bigg| ds \bigg] \]
\[
\leq C \int_0^T e^{-c_s \left( \mathbb{E} \left[ (1 + |\tilde{Y}_{t}^{1,N}_{T-\eta(s)}|^m) \right] \right)^{1/2}} \times \left( \mathbb{E} \left[ \left( \int_{\eta(s)}^{s} \langle \partial_x b(\tilde{Y}_{t}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu), b_t(\tilde{Y}_{t}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) \rangle + \frac{1}{2} \text{tr}\left( \sigma(\tilde{Y}_{t}^{1,N}, \nu) \partial_x \partial_x b(\tilde{Y}_{t}^{1,N}, \mathcal{L}(X_{\eta(s)}), \nu) \sigma^*(\tilde{Y}_{t}^{1,N}, \nu) \right) dt \right) \right) \bigg| ds \bigg] \]
\[
\leq C \int_0^T e^{-c_s \left( \mathbb{E} \left[ (1 + |\tilde{Y}_{t}^{1,N}_{T-\eta(s)}|^m) \right] \right)^{1/2}} \left( \mathbb{E} \left[ \left( \int_{\eta(s)}^{s} (1 + |\tilde{Y}_{t}^{1,N}|^{p+2} + \mathbb{E}|X_{\eta(s)}|^2) dt \right) \right) \bigg| ds \bigg]^{1/2} \]
\[
\leq C h.
\]

Again, using exponential time decay of \( \partial_x V \), in addition to the Lipschitz in measure behaviour of \( b \), uniform moments bounds and crucially the assumption of linearity of the drift in the first component,

\[
A_4 = \left| \mathbb{E} \left[ \int_0^T \left( \partial_x V(s, \tilde{Y}_{t}^{1,N}_{T-s}), b(\tilde{Y}_{\eta(s)}^{1,N}, \mathbb{E}[\alpha(\tilde{Y}_{\eta(s)}^{1,N})], \nu) - b(\tilde{Y}_{\eta(s)}^{1,N}, \mathbb{E}[\alpha(\tilde{Y}_{\eta(s)}^{1,N})], \nu) \right) \right] \right| \]
\[
\leq C \int_0^T e^{-c_s \left( \mathbb{E} \left[ (1 + |\tilde{Y}_{t}^{1,N}_{T-s})| \right] \mathbb{E}[\alpha(\tilde{Y}_{\eta(s)}^{1,N})] - \mathbb{E}[\alpha(\tilde{Y}_{\eta(s)}^{1,N})] \right) \bigg| ds \bigg] \]
\[
\leq C \int_0^T e^{-c_s \left( \mathbb{E} \left[ (1 + |\tilde{Y}_{t}^{1,N}_{T-s})| \right] \mathbb{E}[\alpha(\tilde{Y}_{\eta(s)}^{1,N})] - \mathbb{E}[\alpha(\tilde{Y}_{\eta(s)}^{1,N})] \right) \bigg| ds \bigg] \]
\[
\leq C \int_0^T e^{-c_s \left( \mathbb{E} \left[ |\alpha(\tilde{Y}_{\eta(s)}^{1,N})| \right] - \frac{1}{N} \sum_{j=1}^{N} \alpha(\tilde{Y}_{\eta(s)}^{j,N}) \right) \bigg| ds \bigg] \]
\[
\leq C \int_0^T \mathbb{E} \left[ |\partial_x V(s, \tilde{Y}_{t}^{1,N}_{T-s})| \mathbb{E}[\alpha(\tilde{Y}_{\eta(s)}^{1,N})] - \frac{1}{N} \sum_{j=1}^{N} \alpha(\tilde{Y}_{\eta(s)}^{j,N}) \right] \bigg| ds \bigg] \]
\[
\times \left| \int_0^1 \partial_y b(\tilde{Y}_{\eta(s)}^{Y_{\eta(s)}^{1,N}}, \nu \mathbb{E}[\alpha(\tilde{Y}_{\eta(s)}^{1,N})] + (1 - v) \frac{1}{N} \sum_{j=1}^{N} \alpha(\tilde{Y}_{\eta(s)}^{j,N}), \nu) - \partial_y b(\tilde{Y}_{\eta(s)}^{Y_{\eta(s)}^{1,N}}, \mathbb{E}[\alpha(\tilde{Y}_{\eta(s)}^{1,N})], \nu) \right| \bigg| ds \bigg] \]
\[
+ C \int_0^T \mathbb{E} \left[ \left( \partial_x V(s, \tilde{Y}_{t}^{1,N}_{T-s}), \left( \mathbb{E}[\alpha(\tilde{Y}_{\eta(s)}^{1,N})] - \frac{1}{N} \sum_{j=1}^{N} \alpha(\tilde{Y}_{\eta(s)}^{j,N}) \right) \partial_y b(\tilde{Y}_{\eta(s)}^{Y_{\eta(s)}^{1,N}}, \mathbb{E}[\alpha(\tilde{Y}_{\eta(s)}^{1,N})], \nu) \right) \right] \bigg| ds \bigg] \]
\[
(C.5.10)
\]
\[
(C.5.11)
\]

For simplicity let us name the two previous terms as \( A_6 := (C.5.10) \) and \( A_7 := (C.5.11) \) respectively. Using the Lipschitz condition on \( \partial_y b \), Cauchy–Schwarz inequality, the decay of \( \partial_x V \),
and uniform bounds of the moments, $A_6$ can be bounded as follows:

\[
A_6 = \int_0^T \mathbb{E} \left[ \left| \partial_x V(s, \bar{Y}^{1,N}_{T-s}) \right| \left| \mathbb{E} \left[ \alpha(\bar{Y}^{1,N}_s) \right] - \frac{1}{N} \sum_{j=1}^N \alpha(\bar{Y}^{j,N}_{\eta(s)}) \right| \times \right. \\
\left. \times \left. \int_0^1 \partial_y b(\bar{Y}^{1,N}_{\eta(s)}, \nu \mathbb{E}[\alpha(\bar{Y}^{1,N}_{\eta(s)})] + (1 - \nu) \frac{1}{N} \sum_{j=1}^N \alpha(\bar{Y}^{j,N}_{\eta(s)}), \nu \right) \\
- \partial_y b(\bar{Y}^{1,N}_{\eta(s)}, \mathbb{E}[\alpha(\bar{Y}^{1,N}_{\eta(s)})], \nu) \right] ds \\
\leq C \int_0^T \mathbb{E} \left[ \left| \partial_x V(s, \bar{Y}^{1,N}_{T-s}) \right| \left| \mathbb{E}[\alpha(\bar{Y}^{1,N}_s)] \right| \\
- \frac{1}{N} \sum_{j=1}^N \alpha(\bar{Y}^{j,N}_{\eta(s)}) \left| \int_0^1 (1 - \nu) \mathbb{E}[\alpha(\bar{Y}^{1,N}_s)] - \frac{1}{N} \sum_{j=1}^N \alpha(\bar{Y}^{j,N}_{\eta(s)}) \right| d\nu \right] ds \\
\leq C \int_0^T e^{-cs} \left( \mathbb{E} \left[ \left( 1 + |\bar{Y}^{1,N}_{T-s} | \right)^m \right] \right)^{1/2} \left\{ \mathbb{E} \left[ \left( \left| \frac{1}{N} \sum_{j=1}^N \alpha(\bar{Y}^{j,N}_{\eta(s)}) - \mathbb{E}[\alpha(\bar{Y}^{1,N}_s)] \right|^{4} \right) \right] \right\}^{1/2} ds \\
\leq C \int_0^T e^{-cs} \left( \mathbb{E} \left[ \left| \frac{1}{N} \sum_{j=1}^N \alpha(\bar{Y}^{j,N}_{\eta(s)}) - \mathbb{E}[\alpha(\bar{Y}^{1,N}_s)] \right|^{4} \right] \right) \frac{1}{2} ds.
\]

Now, consider $\bar{X}^{j,N}_{\eta(s)}$ i.i.d copies of the process $(\bar{X}_t)_{t \geq 0}$ evaluated at the node $\eta(s)$. We can then recast the above into two forth–order moments of the centred expressions of independent and i.i.d random variables in the following way:

\[
C \int_0^T e^{-cs} \left( \mathbb{E} \left[ \left| \frac{1}{N} \sum_{j=1}^N \alpha(\bar{Y}^{j,N}_{\eta(s)}) - \mathbb{E}[\alpha(\bar{Y}^{1,N}_s)] \right|^{4} \right] \right)^{1/2} ds
\]

\[
\leq C \int_0^T e^{-cs} \left( \mathbb{E} \left[ \left| \frac{1}{N} \sum_{j=1}^N \alpha(\bar{Y}^{j,N}_{\eta(s)}) - \mathbb{E}[\alpha(\bar{Y}^{1,N}_s)] \right|^{4} \right] \right)^{1/2} ds.
\]

By [93, Corollary 5.12] and uniform strong propagation of chaos ( [50, Theorem 2.4]), we have that both terms are of order $1/N$, and consequently, $A_6 \leq C/N$, where the constant again does not depend on $N$ or $T$.

On the other hand, for the term $A_7$, notice that given the symmetric nature of the problem, i.e. $\bar{Y}^{j,N}$ are exchangeable and i.i.d, we can substitute the term

\[
\mathbb{E} \left[ \langle \partial_x V(s, \bar{Y}^{1,N}_{T-s}), \left( \mathbb{E}[\alpha(\bar{Y}^{1,N}_s)] - \frac{1}{N} \sum_{j=1}^N \alpha(\bar{Y}^{j,N}_{\eta(s)}) \right) \partial_y b(\bar{Y}^{1,N}_{\eta(s)}, \mathbb{E}[\alpha(\bar{Y}^{1,N}_s)], \nu) \right] \\
\leq \mathbb{E} \left[ \left( \frac{1}{N} \sum_{i=1}^N \partial_x V(s, \bar{Y}^{i,N}_{T-s}), \left( \mathbb{E}[\alpha(\bar{Y}^{i,N}_s)] - \frac{1}{N} \sum_{j=1}^N \alpha(\bar{Y}^{j,N}_{\eta(s)}) \right) \partial_y b(\bar{Y}^{i,N}_{\eta(s)}, \mathbb{E}[\alpha(\bar{Y}^{i,N}_s)], \nu) \right] \\
\leq \mathbb{E} \left[ \left( \frac{1}{N} \sum_{i=1}^N \partial_x V(s, \bar{Y}^{i,N}_{T-s}), \left( \mathbb{E}[\alpha(\bar{Y}^{i,N}_s)] - \frac{1}{N} \sum_{j=1}^N \alpha(\bar{Y}^{j,N}_{\eta(s)}) \right) \partial_y b(\bar{Y}^{i,N}_{\eta(s)}, \mathbb{E}[\alpha(\bar{Y}^{i,N}_s)], \nu) \right] \right].
\]
Moreover, since $\mathbb{E}[\alpha(\bar{Y}_{\eta(s)}^{1,N})] = \frac{1}{N} \sum_{j=1}^{N} \alpha(\bar{Y}_{\eta(s)}^{j,N})$ is a centred expression, we can always subtract its expectation, which is 0, and further substitute (C.5.13) by

$$
\mathbb{E}\left[\frac{1}{N} \sum_{i=1}^{N} \partial_{x} V(s, \bar{Y}_{T-s}^{i,N}), \left\{ \mathbb{E}[\alpha(\bar{Y}_{\eta(s)}^{i,N})] - \frac{1}{N} \sum_{j=1}^{N} \alpha(\bar{Y}_{\eta(s)}^{j,N}) \right\} \partial_{y} b(\bar{Y}_{\eta(s)}^{i,N}, \mathbb{E}[\alpha(\bar{Y}_{\eta(s)}^{i,N})], \nu) \right]
$$

$$
- \mathbb{E}\left[\partial_{x} V(s, \bar{Y}_{T-s}^{1,N}), \left\{ \mathbb{E}[\alpha(\bar{Y}_{\eta(s)}^{1,N})] - \frac{1}{N} \sum_{j=1}^{N} \alpha(\bar{Y}_{\eta(s)}^{j,N}) \right\} \partial_{y} b(\bar{Y}_{\eta(s)}^{1,N}, \mathbb{E}[\alpha(\bar{Y}_{\eta(s)}^{1,N})], \nu) \right] \right].
$$

Finally, again by the the fact that for all $j \in \{1, \ldots, N\}$ we have $\mathbb{E}[\alpha(\bar{Y}_{\eta(s)}^{j,N})] = \mathbb{E}[\alpha(\bar{Y}_{\eta(s)}^{1,N})]$, the above expression is equal to

$$
\mathbb{E}\left[\frac{1}{N} \sum_{i=1}^{N} \partial_{x} V(s, \bar{Y}_{T-1-s}^{i,N}), \left\{ \mathbb{E}[\alpha(\bar{Y}_{\eta(s)}^{i,N})] - \frac{1}{N} \sum_{j=1}^{N} \alpha(\bar{Y}_{\eta(s)}^{j,N}) \right\} \partial_{y} b(\bar{Y}_{\eta(s)}^{i,N}, \mathbb{E}[\alpha(\bar{Y}_{\eta(s)}^{i,N})], \nu) \right]
$$

$$
- \mathbb{E}\left[\partial_{x} V(s, \bar{Y}_{T-s}^{1,N}), \left\{ \mathbb{E}[\alpha(\bar{Y}_{\eta(s)}^{1,N})] - \frac{1}{N} \sum_{j=1}^{N} \alpha(\bar{Y}_{\eta(s)}^{j,N}) \right\} \partial_{y} b(\bar{Y}_{\eta(s)}^{1,N}, \mathbb{E}[\alpha(\bar{Y}_{\eta(s)}^{1,N})], \nu) \right] \right].
$$

As a consequence, after applying twice the Cauchy–Schwarz inequality and the same result [93, Corollary 5.12] used identically to (C.5.12), $A_{7}$ can be bounded as it follows:

$$
A_{7} \leq C \int_{0}^{T} e^{-c_{s}} \left( \mathbb{E}\left[ \left| \frac{1}{N} \sum_{j=1}^{N} \alpha(\bar{Y}_{\eta(s)}^{j,N}) - \mathbb{E}[\alpha(\bar{Y}_{\eta(s)}^{1,N})] \right|^{4} \right] \right)^{1/4} \times
$$

$$
\times \left( \mathbb{E}\left[ \left| \frac{1}{N} \sum_{i=1}^{N} \partial_{x} V(s, \bar{Y}_{T-s}^{i,N}) \partial_{y} b(\bar{Y}_{\eta(s)}^{i,N}, \mathbb{E}[\alpha(\bar{Y}_{\eta(s)}^{i,N})], \nu) \right|^{4} \right] \right)^{1/4} - \mathbb{E}\left[ \partial_{x} V(s, \bar{Y}_{T-s}^{1,N}) \partial_{y} b(\bar{Y}_{\eta(s)}^{1,N}, \mathbb{E}[\alpha(\bar{Y}_{\eta(s)}^{1,N})], \nu) \right|^{4} \right]^{1/4}
$$

$$
\leq C \frac{1}{\sqrt{N}} \frac{1}{\sqrt{N}}.
$$

To conclude, we have proven that

$$
|\mathbb{E}[\phi(\bar{Y}_{T-s}^{1,N})] - \mathbb{E}[\phi(X_{T})]| \leq C \left( h + \int_{0}^{T} e^{-c_{s}} \mathbb{E}\left[ \left| \mathbb{E}[\alpha(X_{s})] - \mathbb{E}[\alpha(\bar{Y}_{\eta(s)}^{1,N})] \right| ds + \frac{1}{N} \right] \right)
$$

(C.5.14)

holds for an arbitrary measurable $\phi$. If now $\phi$ is taken to be precisely $\alpha$ (notice we must either have $\alpha : \mathbb{R}^{d} \to \mathbb{R}$ or $\phi : \mathbb{R}^{d} \to \mathbb{R}^{d}$, in which case all the computations in the previous sections are done component-wise for $\partial_{x} V^{i}$ but denoted as $\partial_{x} V$), (C.5.14) implies by Grönwall’s lemma that there exists $C > 0$ such that

$$
|\mathbb{E}[\alpha(X_{t})] - \mathbb{E}[\alpha(\bar{Y}_{\eta(t)}^{1,N})]| \leq C(h + 1/N), \quad \forall t \in [0, T].
$$

Moreover, on its turn when this bound is taken back to (C.5.14), we obtain that the weak error itself is of order $(h + 1/N)$ uniformly in time. \qed


167