On the McKean-Vlasov Dynamics with or without Common Noise

William Richard Philip Hammersley

Submitted for the Degree of
Doctor of Philosophy

2020

The University of Edinburgh

Abstract

McKean-Vlasov stochastic differential equations may arise from a probabilistic interpretation of certain non-linear PDEs or as the limiting behaviour of mean field particle systems (those whose interactions are through the empirical measure) as the population size increases to infinity. Interest in this topic has grown enormously in recent times following the introduction of the related mean field games. These are models derived from the infinite population limit of games with finitely many players and mean field structure, i.e. the dynamics and rewards of one player depend on the other players through the empirical measure. Naturally, it is imperative that the dynamics of the models are well-posed. This question comprises the majority of this text in two stochastic contexts: with or without a common noise.

In the more often studied case where the particles are driven by independent Brownian motions, results are provided that pertain to the weak-existence and pathwise continuous dependence on the initial condition. These results adapt a method of Gyöngy and Krylov for Itô's stochastic differential equations to the McKean-Vlasov setting. Should the coefficients and initial distribution satisfy a certain Lyapunov condition, well-posedness of the dynamics may be established along with the existence of an invariant measure for an associated semi-group. These conditions allow for potentially unbounded coefficients, with growth intrinsically linked to the Lyapunov condition.

In the second context, particle systems driven by correlated noises are considered. In particular, the particles are each driven by two Brownian motions: one common to all particles and a private Brownian motion independent of all others. The connection between these particle systems and related McKean-Vlasov models through the conditional propagation of chaos is discussed. Existence and uniqueness of weak solutions to the corresponding McKean-Vlasov dynamics is proved in a particular framework that allows for a discontinuous drift coefficient at a price of non-degenerate noise.

Lay Summary

Systems are ubiquitous. At every level of our world, from the natural to the sociological and technological, complex and networked activity is observed. The study of mean field models provides an approach to develop our understanding of systems where the number of interacting elements is large and the interactions have a particular structure, by considering a corresponding system of infinite size. The entities that comprise a system may be, for example, molecules, animals, people, etc. Consequently, the scope for application is enormous.

The main contribution of this text is to broaden the collection of systems for which there exist unique solutions to the equations governing the infinite system.

Declaration

I declare that the thesis has been composed by myself and the work therein is the product of my own efforts and collaboration with my supervisors, Łukasz Szpruch and David Šiška. Chapters 2 and 4 and their corresponding appendices are to be published in Annales de l'Institut Henri Poincaré (B) and the Annals of Probability, respectively. The right to use the material contained in those chapters, here and in other future works, was reserved. This work has not been submitted for any other degree of professional qualification.

(William Hammersley)

Acknowledgements

Whilst at times the PhD has felt like a long and lonely road, I count myself extremely lucky to have been surrounded by so many kind and creative people who have guided, motivated, and inspired me to reach this point.

Firstly, I would like to express my gratitude to my supervisors Łukasz Szpruch and David Šiška for introducing me to this area of Mathematics and showing me the ropes of research. The confidence and courage they have given me to pursue my own ideas has been instrumental in my mathematical and personal development. Their pragmatic advice has proved invaluable during the PhD and for this I am truly grateful.

Naturally, I am indebted to the Mathematics community at large. However, I would particularly like to acknowledge the help I received from Sandy Davie and Daniel Lacker. Their generous input helped me overcome hurdles at various stages of my research. I would also like to thank Xīlíng Zhāng and Jean-François Jabir for their advice and conversation. Additionally, I would like to thank the anonymous referees for their time and effort given in reviewing the papers produced.

On a personal note, my time in Edinburgh has been unexpectedly rich and varied. I would like to thank the friends I have made here for all the good times we have shared and the support they have given.

Finally, I would like to thank my family. Their encouragement throughout my education has been unwavering.

To Eléonore

Contents

Chapter 1

Introduction

Systems are ubiquitous. At every level of our world, from the natural to the sociological and technological, complex and networked activity is observed. The study of mean field models provides an approach to develop our understanding of systems where the number of interacting elements is large and the interactions have a particular structure.

In reality, the modelled systems have a finite population of interacting elements. Unfortunately, these models may be difficult to analyse and to approximate numerically due to the sheer number of interactions. This intractability is an instance of the curse of dimensionality. In a particular class of large population models - those in which interaction is through the empirical measure - one is able to simplify the analysis to some extent, by approximating macroscopic properties of the population or those of a representative particle with the mean field limit: the limiting behaviour as the population of interacting elements increases to infinity.

The mechanism by which this approach becomes possible is referred to as the propagation of chaos. Starting from a 'chaotic' (independent) initial distribution of particles and driven by independent noises, finite mean field particle systems become correlated through their interactions. However, courtesy of the mean field structure of the system, it may be shown that sub-populations of fixed size from a sequence mean field particle systems (increasing in population) becomes decoupled in the infinite population limit. In the infinite system, the initial chaos propagates through time. The now independent and identically distributed individuals from the infinite system have time-marginals that are governed by a non-linear evolution; this is referred to as the McKean-Vlasov equation.

Mean field models have a broad range of applications; the constituent elements of a system may be for example, people, animals, cells, molecules or algorithmic parameters. In many cases, to formulate a realistic model, it is necessary to account for the agency of individuals. These models fall under the term mean field games.

Since its 2006 début [47, 48, 72, 73, 74], the theory of mean field games has garnered a huge amount of attention, along with the related class of mean field control problems where the populations' actions are dictated by a social planner. Whilst the structure of these models is quite particular, it is worth remarking that many systems fit within this category. See the two volume text [26] for a general overview of the field from a stochastic perspective.

Introducing mean field models as approximations to finite mean field particle systems is however, a more modern perspective. At its inception, the study of finite mean field particle systems and their corresponding mean field limits was motivated as a probabilistic method of constructing solutions to certain non-linear partial differential equations. From this converse perspective, the infinite system is the object of interest, rather than the finite particle systems. The study of mean field particle systems from this original viewpoint was initiated by McKean and Kac. In [80], McKean introduced a class of Markov processes whose time marginals satisfy certain non-linear equations. This elucidated the wider consequences of Kac's probabilistic interpretation of Maxwell's assumption of molecular chaos in the kinetic theory of gases [54]. Further, in [81], McKean demonstrated how to construct solutions from a sequence of interacting particle systems of increasing population size via the aforementioned propagation of chaos. Since these pioneering efforts, the theory and application of McKean-Vlasov models has flourished, permeating many fields of study.

Fundamental to these models is the underlying question of well-posedness of the dynamics. This is the main focus of the thesis.

1.1 Overview of the Thesis

To give a sense of this text as a whole, numerous results contained in this thesis and the challenges that were overcome to produce them are described in this section. For reviews of the related literature, the reader is referred to the corresponding chapters.

In the situation where the particles are driven by independent noises, well-posedness of the McKean-Vlasov stochastic differential equation is established, following an approach inspired by that of Gyöngy and Krylov [41] for classical Itô SDEs. In the presence of certain Lyapunov criteria - a condition dependent on the structure of the coefficients and the initial condition - one may obtain estimates ensuring the non-explosion of approximate solutions, enabling the use of a localisation procedure. In the classical case of Itô's SDEs a critical fact is, that if two (locally) uniquely solvable SDEs started from the same point have the same coefficients on some domain, then up until the first exit time of the domain, both solutions coincide. This statement no longer holds true in the McKean-Vlasov setting as the coefficients depend on the law of the solution; the McKean-Vlasov SDE is non-local. It is easy to see that, in general, the law of a stopped process will not be equal to the law of the un-stopped process. If one attempts a naïve extension of Gyöngy and Krylov's method, one would be tempted to formulate rather difficult-to-verify conditions ensuring that the estimates one searches for are attainable. This would be problematic since Lyapunov criteria can, at times, be difficult to find. However, changing the perspective from stopping a process once it exits a domain, to truncating its coefficients outside the domain, one obtains a useful fact; the law of a solution to the truncated SDE is equal to the law of that same solution stopped at the exit of the domain. This fact enables more readily available conditions to be formulated.

By use of the differential calculus of Lions [78], one can consider Lyapunov functions taking arguments in the space of probability measures. This is a natural extension of existing methods since the coefficients themselves also have probability measure valued arguments. After formulating appropriate Lyapunov conditions and related growth assumptions for the coefficients, the existence of solutions follows from careful application of Skorokhod's method for constructing solutions to SDEs [91] and a localisation procedure. A two-fold approximation is made; the coefficients are truncated on an increasing, nested sequence of sub-domains and suitable Euler schemes are used to approximate a solution to the truncated SDE. Regarding the continuous dependence on initial conditions of solutions, criteria are introduced that extend the usual monotonicity condition. This narrows a gap in the literature, as there are examples of non-uniqueness of McKean-Vlasov SDEs in the locally-Lipschitz regime [90]. Finally, the existence of an invariant measure for a semi-group associated with the flow of marginal distributions of solutions to the McKean-Vlasov SDE is established via the Krylov-Bogolyubov theorem [63].

In the setting with a common noise, existence and uniqueness of weak solutions is proved for a framework inspired by that of Mishura and Veretennikov [84] for the case without common noise. The driving noises of the particle systems from which the McKean-Vlasov SDE with common noise is derived are correlated, resulting in a more intricate model. The coefficients of this limiting equation depend on the conditional law of a solution given a sub-filtration to which the common noise is adapted. In order to both proceed with the standard compactness methods to obtain weak solutions and to connect these solutions to an associated stochastic partial differential equation, one needs to carefully define the compatibility structure (see Kurtz [65]). This thesis sheds light on such a definition from both the above perspective and also via the conditional propagation of chaos. With definition in hand, weak existence is demonstrated first under modest assumptions of bounded and continuous coefficients via Euler-type approximations. Then, the existence of solutions is proved for a particular class of Markovian coefficients, where the dependence on measure of the coefficients is via integration of so called interaction kernels. At a price of non-degeneracy, existence may be established in the case where the interaction kernels are only assumed to be bounded and measurable, by use of Krylov's estimates [64].

The uniqueness in joint law of solutions to the McKean-Vlasov with common noise is proved via an extension of the method of Mishura and Veretennikov [84] for weak uniqueness in the case without a common noise. Due to the dependence structure of solutions, in this context it is natural to consider joint weak uniqueness rather than the uniqueness in law. Mishura and Veretennikov's approach is closely linked to Pinsker's inequality, where one estimates the square of total variation by half the Kullback-Leibler divergence. By representing the two solutions via Girsanov transformations from an intermediary probability space, they are able to prove weak uniqueness if the measure dependence of the coefficients is only in the drift and is total-variation Lipschitz. Extending this argument to the common noise setting presents some challenges. Firstly, one starts with a probability space supporting both weak solutions. The ability to represent two solutions (without common noise) via Girsanov transformations relies on the uniqueness in law of solutions to the Itô SDEs obtained by fixing the coefficients with the law of each solution. In the common noise setting, the comparable argument of fixing the flow of conditional distributions of the solutions required the uniqueness in law of solutions to the SDE with random drift coefficient to be established. Furthermore, the Girsanov transformation would need to leave unaffected, the distributions of processes adapted to the conditioning sub-filtration. This requirement is fulfilled courtesy of the dependence structure demanded of solutions. Secondly, since the flow of conditional distributions given the common noise is stochastic, when one wishes to estimate the total variation between them, it helps for them to have the same conditioning sigma algebras. To understand why, consider the case where one has a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, one which there are defined two random elements X^1 and X^2, and a sub-σ-algebra $\mathcal{G}$. Then, letting $\mathscr{L}(X^i|\mathcal{G})$ denote the regular conditional distribution of X^i given $\mathcal{G}$, for $i = 1, 2$, one has the following estimate:

$$\mathbb{E}[d_{TV}(\mathscr{L}(X^1|\mathcal{G}), \mathscr{L}(X^2|\mathcal{G}))] \leq \mathbb{E}[\mathbb{E}[\mathbb{1}_{X^1 \neq X^2}|\mathcal{G}]] = \mathbb{E}[\mathbb{1}_{X^1 \neq X^2}],$$

where d_{TV} is the total variation distance. Let $\mathscr{L}(X^i)$ denote the law of X^i, for $i = 1, 2$ and recall that there exists an optimal coupling $\mathbb{P}^*$ for which $d_{TV}(\mathscr{L}(X^1), \mathscr{L}(X^2)) = \mathbb{E}^*[\mathbb{1}_{X^1 \neq X^2}]$. Suppose $\mathbb{P} = \mathbb{P}^*$, and that $d_{TV}(\mathscr{L}(X^1), \mathscr{L}(X^2))$ could be estimated by $\alpha\mathbb{E}[d_{TV}(\mathscr{L}(X^1|\mathcal{G}), \mathscr{L}(X^2|\mathcal{G}))]$ (for $\alpha < 1$), then one could conclude that X^1 and X^2 have the same distribution. This is analogous to the situation one would hope to be in to proceed with the argument of [84]. However, there is no guarantee a priori that the optimal coupling for the total variation between two solutions to the McKean-Vlasov with common noise, should constrain the common Brownian motion and the flow of conditional measures to generate the same sub-filtrations. To counteract this issue, rather than using total variation, the cost function in the Kantorovich problem is altered so that its optimal coupling precludes discrepancies in the randomness that generates the conditioning sub-filtrations. Choosing the original probability space from which the Girsanov transformations are to be defined to be the optimal coupling of this new Kantorovich problem, a similar estimation procedure to that of [84] may be followed to prove joint weak uniqueness.

1.1.1 Structure of the Thesis

The rest of the current chapter provides an introduction to mean field particle systems and the corresponding McKean-Vlasov equations. In the interests of clarity and concision, the models of mean field games and control are not introduced as the thesis does not require any knowledge of them. Chapter 1 concerns the well-posedness of the McKean-Vlasov stochastic differential equation in the absence of a common noise and the existence of an invariant measure for its associated semi-group. Chapter 2 expounds on the mean field limit in the context of common noise via the conditional propagation of chaos. Chapter 3 concerns weak existence and uniqueness for the McKean-Vlasov stochastic differential equation with a common noise. Finally, a short discourse on potential extensions of the work contained in this thesis is provided. This thesis contains material from two articles [44] and [45] that are to appear in Annales de l'Institute Henri Poincaré (B) Probabilités et Statistiques and the Annals of Probability, respectively. As such, these chapters are relatively self-contained, however there are some repeated statements and slight differences of notational choice between Chapter 2 and the rest of the thesis.

1.2 Notation

- The law of a random element, say Z, will be denoted $\mathscr{L}(Z)$.

- Time indexes are denoted by I and will be either $[0, T]$ for some $T < \infty$ or $\mathbb{R}^+ := [0, \infty)$ dependent on context.

- For a topological space E, the Borel sigma algebra on E is denoted $\mathcal{B}(E)$.

- Given a stochastic process X and a stopping time τ, the process X stopped at time τ will be denoted $X_{\cdot\wedge\tau} := \{X_{t\wedge\tau}\}_{t\in I}$. Let the filtration generated by X be denoted as $\mathbb{F}^X := \{\mathcal{F}^X_t\}_{t\in I}$. Processes denoted by Roman letters, will be real-vector valued.

- The space of continuous paths from I into $\mathbb{R}^d$ is denoted $\mathcal{C} := C(I; \mathbb{R}^d)$ and will be equipped with either the uniform or Skorokhod metric, depending on context. Let $\mathcal{P}(\mathcal{C})$ denote the set of Borel probability measures on $\mathcal{C}$.

- Coefficients of stochastic differential equations will usually be denoted b, σ and ρ. They are assumed to be (or are at least identified with) Borel measurable functions from (possibly a subset of) $I \times \mathcal{C} \times \mathcal{P}(\mathcal{C})$ into $\mathbb{R}^d, \mathbb{R}^{d\times d_1}$ and $\mathbb{R}^{d\times d_2}$, respectively (dimensions will be determined by context), that are always assumed to be at least progressive. To clarify, a function f on $I \times \mathcal{C} \times \mathcal{P}(\mathcal{C})$ is called progressive if for any $t \in I$,

$$f(t, x, m) = f(t, x_{\cdot\wedge t}, m \circ \phi_t^{-1}), \text{ where } \phi_t : \mathcal{C} \ni x \mapsto x_{\cdot\wedge t} \in \mathcal{C}.$$

  A function f on $I \times \mathcal{C} \times \mathcal{P}(\mathcal{C})$ is called Markovian if for any $t \in I$,

$$f(t, x, m) = \tilde{f}(t, x_t, m \circ \pi_t^{-1}), \text{ where } \pi_t : \mathcal{C} \ni x \mapsto x_t \in \mathbb{R}^d,$$

  for a measurable function $\tilde{f}$ with domain $I \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$. In the case of Markovian coefficients, b, σ and ρ may instead be defined as their counterparts with a tilde.

- The set of positive integers are denoted $\mathbb{N}$ and the non-negative integers as $\mathbb{N}_0$.

- Given a measure space $(\Omega, \mathcal{F}, \mu)$ and an integrable function f, let

$$\langle \mu, f \rangle := \int_\Omega f \, d\mu.$$

- The transpose of a matrix a will be written as a^T and the trace of a square matrix is denoted $\mathrm{tr}(a)$. For two real valued column vectors, a and b, the inner product $a^T b$ may be written as ab.

- Let the gradient be denoted as ∂_x and the Hessian as ∂_x^2.

- All stochastic integrals in this thesis are Itô integrals.

1.3 A Mean Field Particle System

An N-particle mean field system $(X^{1,N}, \dots, X^{N,N})$, may satisfy the following stochastic differential equations, for $i = 1, \dots, N$,

$$\begin{aligned} X_t^{i,N} &= X_0^i + \int_0^t b(s, X^{i,N}, \mu^N)ds + \int_0^t \sigma(s, X^{i,N}, \mu^N)dW_s^i, \\ \mu^N &:= \frac{1}{N}\sum_{j=1}^N \delta_{X^{j,N}}, \end{aligned} \tag{MFS}$$

with Brownian motions W^i and initial values X_0^i. The standard mean field models assume independence of the initial conditions and of the driving Brownian motions. A crucial observation is that under these conditions (and those ensuring weak uniqueness [98]), the particle system is exchangeable, i.e. the distribution of $(X^{1,N}, \dots X^{N,N})$ is invariant under permutation of the indices. As an extension to the standard mean field models, one may consider cases where the inputs X_0^i and W^i are correlated,[1] introducing input noise common to all particles. These more complicated models are the subject of the latter half of the thesis and their treatment is postponed.

1.4 The Mean Field Limit

Assume momentarily that as the number of particles N increases, μ^N has a deterministic limit μ and the particles $X^{i,N}$ converge pointwise to processes $X^{i,\infty}$, satisfying the equations

$$X_t^{i,N} = X_0^i + \int_0^t b(s, X^{i,N}, \mu)ds + \int_0^t \sigma(s, X^{i,N}, \mu)dW_s^i$$

These limit processes (at least in the strong solution setting) are independent, by the independence assumption on the stochastic inputs X_0^i and W^i. For continuous and bounded f, $\mathbb{E}[f(X^{i,N})] \to \mathbb{E}[f(X^{i,\infty})]$ and, due to the exchangeability of the particle systems,

$$\mathbb{E}[f(X^{i,N})] = \mathbb{E}[\langle \mu^N, f\rangle] \to \langle \mu, f\rangle.$$

Therefore, one expects that a selection of particles from the mean field systems should converge to independent solutions of the following McKean-Vlasov stochastic differential equation,

$$\begin{aligned} X_t &= X_0 + \int_0^t b(s, X, \mu)ds + \int_0^t \sigma(s, X, \mu)dW_s, \\ \mu &= \mathscr{L}(X). \end{aligned} \tag{MKV}$$

Equivalent characterisations of this candidate mean field limit are available under certain assumptions as solutions to an associated non-linear Fokker-Planck-Kolmogorov

[1] A further extension may also be considered, where a small number of particles are distinguished. In the game theoretic setting, these are mean field games with *major players* [25, 46].

equation or as solutions to a non-linear martingale problem. The former is called the McKean-Vlasov equation and is defined via the action on test functions as

$$\langle \mu_t, \phi \rangle = \langle \mu_0, \phi \rangle + \int_0^t \langle \mu_s, b(s,\cdot,\mu)\partial_x f(\pi_s(\cdot)) + \frac{1}{2} tr(\sigma\sigma^T(s,\cdot,\mu)\partial_x^2 f(\pi_s(\cdot)))\rangle ds \tag{FPK}$$

where $\phi \in C_0^2$ and π_s are projections from $\mathcal{C}$ into $\mathbb{R}^d$ such that $\mathcal{C} \ni x \mapsto x_s \in \mathbb{R}^d$. That such a characterisation is equivalent (subject to assumptions) to a solution of the McKean-Vlasov stochastic differential equation is due to the so called superposition principle - see [37], the extension [97] and for superposition principles in the common noise setting, [71]. A simpler equivalence of characterisation is available (again, subject to assumptions) through an associated non-linear martingale problem:

Definition 1.4.1. Let X be the canonical process on $\mathcal{C}$. A probability measure μ on $\mathcal{C}$ is a solution to the non-linear martingale problem if, for every $\phi \in C_0^2$,

$$\phi(X_t) - \phi(X_0) - \int_0^t b(s,X,\mu)\partial_x f(X_s) + \frac{1}{2} tr(\sigma\sigma^T(s,X,\mu)\partial_x^2 f(X_s)) \tag{1.4.1}$$

is a martingale under μ with respect to the natural filtration.

That a solution to the non-linear martingale problem gives rise to a solution of the McKean-Vlasov stochastic differential equation follows from an analogous procedure to the classical case - see for example Proposition 5.4.6 in [57] for the classical case and Proposition 3.1 in [77] for the McKean-Vlasov case.

1.5 The Propagation of Chaos

An equivalence given in Sznitman's 'topics in propagation of chaos' [93], clarifies the connection between finite mean field particle systems and the above mean field limits. The following definition and proposition are taken from [93].

Definition 1.5.1. Let E be a separable metric space, and $\{u_N\}_{N\in\mathbb{N}}$ a sequence of symmetric probability measures, each defined on a corresponding product space E^N. Say that u_N is u-chaotic, for u a probability measure on E, if for $\phi_i \in C_b(E)$, and any $k \in \mathbb{N}$,

$$\lim_{N\to\infty} \langle u_N, \phi_1 \otimes \cdots \otimes \phi_k \otimes 1 \cdots \otimes 1 \rangle = \prod_{i=1}^k \langle u, \phi_i \rangle \tag{1.5.1}$$

Proposition 1.5.2. *i) u_N is u-chaotic is equivalent to $\bar{X}_N = \frac{1}{n}\sum_{i=1}^N \delta_{X^{i,N}}$ ($\mathcal{P}(E)$-valued random elements on (E^N, u_N), $X^{i,N}$ canonical coordinates on E^N) converges in law to the constant random variable u. It is also equivalent to condition (1.5.1) with $k = 2$.*

ii) Let E be a Polish space, and $\{m^n\}_{n\in\mathbb{N}}$ a sequence of probability measures on $\mathcal{P}(E)$. Then tightness of the sequence $\{m^n\}_{n\in\mathbb{N}}$ is equivalent to tightness of the sequence of the intensity measures $I(m^n)$ defined by

$$\langle I(m^n), f \rangle = \int_{\mathcal{P}(E)} \langle \nu, f \rangle m^n(d\nu), \tag{1.5.2}$$

for any bounded measurable f.

Remark 1.5.3. The proposition suggests the following well-known line of argumentation:

1. Establish tightness of the sequence laws of μ^n in $\mathcal{P}(\mathcal{P}(\mathcal{C}))$ or equivalently (by Proposition 1.5.2ii), the tightness of the laws of $\{X^{1,N}\}_{N\in\mathbb{N}}$ in $\mathcal{P}(\mathcal{C})$.

2. Prove that the limits of convergent sub-sequences of $\mathscr{L}(\mu^N)$ are supported on the set of solutions to (FPK) or alternatively, solutions to the non-linear martingale problem of Definition 1.4.1.

3. Establish the uniqueness of solutions to (FPK), the non-linear martingale problem or uniqueness in law for solutions to the McKean-Vlasov SDE (MKV).

Once uniqueness is established, then μ^N converge to a deterministic limit. As a consequence of Proposition 1.5.2, the convergence of the empirical measure μ^N towards a deterministic limit μ is equivalent to the convergence of the induced distributions of a finite collection of k particles from each particle system towards the k-fold product measure $\mu^{\otimes k}$ for any fixed $k \in \mathbb{N}$. In other words, the collection of any k particles converges in distribution to the law of k independent solutions to the McKean-Vlasov stochastic differential equation; in the infinite particle limit, the initial chaos propagates through the system. See [39, 40, 70, 83, 85, 93] for demonstrations of this phenomenon in various contexts.

Alternatively, one may try to prove this convergence directly by constructing a probability space supporting infinitely many independent Brownian motions, each with an associated solution to the McKean-Vlasov stochastic differential equation. Every i^{th} Brownian motion is also the driving Brownian motion of the corresponding i^{th} particle from a sequence of solutions to the particle systems also supported on this probability space. Then, through pathwise considerations, one demonstrates that on this space, collections of k particles from the finite mean field systems converge towards the corresponding independent solutions to the McKean-Vlasov stochastic differential equation, see [83, 93]. The disadvantage of this approach is that it requires stronger assumptions, as is expected when using pathwise techniques. However, it does allow for the quantification of the rate of propagation of chaos. Providing a rate of propagation of chaos is of high practical importance for numerical schemes. There are however, relatively few results in this direction. See for example, [3, 4, 28, 52], for methods of obtaining a rate.

The above discourse highlights the importance of the well-posedness of the McKean-Vlasov limit. As previously mentioned, this question is the dominant topic of the thesis. It is worth mentioning the connection to the propagation of chaos is not the only motivation to study well-posedness of the McKean-Vlasov dynamics. One may be interested only in the non-linear PDE for which there is a stochastic representation through a McKean-Vlasov SDE. In this case, one may wish to study solutions via discretisation schemes rather than considering a finite particle system approximation. This approach is followed in the next chapter where the McKean-Vlasov dynamics are treated in isolation from any related finite particle systems.

1.6 Mean Field Models with Correlated Noise

The type of systemic/common noise considered in this thesis are those cases where the particles have independent initial values and are each driven by independent Brownian motions W^i and a common Brownian motion, B, that is independent of all Brownian motions W^i. The corresponding particle system may be written as,

$$\begin{aligned} X_t^{i,N} &= X_0^{i,N} + \int_0^t b(s, X^{i,N}, \mu^N)\, ds + \int_0^t \sigma(s, X^{i,N}, \mu^N)\, dW_s^{i,N} \\ &\quad + \int_0^t \rho(s, X^{i,N}, \mu^N)\, dB_s, \\ \mu_t^N &:= \frac{1}{N} \sum_{j=1}^N \delta_{X_{\cdot\wedge t}^{j,N}}. \end{aligned} \tag{MFSCN}$$

Since the Brownian motion B is common to all particles, one should not expect the initial independence to propagate in the infinite particle limit. As will be clarified in Chapter 3, the decoupling of the system in this setting is conditional on a filtration to which the systemic noise is adapted. The limiting dynamics for a tuple $(X^{i,N}, \mu^N, W^{i,N}, B)$ will be given as follows,

$$\begin{aligned} X_t &= X_0 + \int_0^t b(s, X, \mu)\, ds + \int_0^t \sigma(s, X, \mu)\, dW_s + \int_0^t \rho(s, X, \mu)\, dB_s, \\ \mu_t &= \mathscr{L}(X_{\cdot\wedge t} | \mathcal{F}_t^{B,\mu}). \end{aligned} \tag{MKVCN}$$

The difficulty of the analysis is elevated in this framework, as the influence of the common noise on the empirical distribution of the particles remains in the infinite particle limit. The now stochastic limit μ for the empirical measures satisfies a fixed point condition that it should be the conditional distribution of the process X given the filtration generated by the Brownian motion B and itself. This additional source of randomness presents interesting challenges even in the case of dynamics, let alone the extended models of control and mean field games, especially when one abandons the Lipschitzian/Monotone setting and considers weak solutions. Chapter 4 provides weak existence and joint uniqueness in law for solutions in a particular framework inspired by that of Mishura and Veretennikov [84].

Connecting solutions to (MFSCN) with solutions to (MKVCN) is the subject of Chapter 3 where a notion of conditional propagation of chaos is introduced. Analogues to the definition of u-chaoticity and the related Proposition (1.5.2) that suit this context are provided, along with conditions under which the conditional propagation of chaos occurs.

Chapter 2

McKean-Vlasov Dynamics without Common Noise: Well-posedness under Lyapunov Conditions

In the standard McKean-Vlasov framework (without a common noise), the results presented in this thesis obtained in collaboration with David Šiška and Łukasz Szpruch pertain to Lyapunov-type criteria that enable access to estimates yielding the existence and continuous dependence on initial conditions of solutions. These results are inspired by the method of Lyapunov as used for Itô's stochastic differential equations by Gyöngy and Krylov [41] and in the McKean-Vlasov setting by Funaki [39] and Gärtner [40] for Lyapunov functions depending on the state. Due to the non-linearity of the McKean-Vlasov equation it is natural to formulate criteria that embrace the measure-dependence of the coefficients. In this vein, Lyapunov functions that may depend upon measure are introduced, requiring use of the Lions/Intrinsic derivative in measure. However, the aforementioned non-linearity introduces subtle, yet critical differences from classical SDE theory, should one wish to depart from the Lipschitzian setting. In particular, the technique of localisation may not be naïvely applied since is it not true a priori that the law of a stopped McKean-Vlasov process is the same as the law of the un-stopped process. Yet with carefully constructed approximations, the rationale of localisation can be followed, resulting in the existence of solutions (See Theorem 2.2.10). In addition to the criteria that enable existence, we introduce conditions extending the usual monotonicity assumption enabling the uniqueness and furthermore the continuous dependence of solutions on initial conditions (see Theorem 2.3.3). As counterexamples to uniqueness of solutions exist for locally Lipschitz coefficients [90], our results narrow a gap in the literature. Finally, under the conditions ensuring existence and uniqueness of solutions, the existence of an invariant measure for an associated semi-group is established

The rest of this chapter and Appendix A are to appear in the Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques under the title 'McKean-Vlasov SDEs under measure dependent Lyapunov conditions'.

2.1 Introduction and Literature Review

This chapter considers either the time interval $I = [0, T]$ for some fixed $T > 0$ or $I = [0, \infty)$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathcal{F}_t)_{t \in I}$ a right continuous filtration such that $\mathcal{F}_0$ contains all sets of $\mathcal{F}$ that have probability zero. Let $w = (w_t)_{t \in I}$ be an $\mathbb{R}^{d'}$-valued an $(\mathcal{F}_t)_{t \in I}$-Wiener process. In this chapter, we consider the McKean-Vlasov stochastic differential equation (SDE) on an open domain $D \subseteq \mathbb{R}^d$ and with Markovian coefficients,

$$x_t = x_0 + \int_0^t b(s, x_s, \mathscr{L}(x_s))\, ds + \int_0^t \sigma(s, x_s, \mathscr{L}(x_s))\, dw_s\,, \quad t \in I\,. \tag{2.1.1}$$

The law of such an SDE satisfies a nonlinear Fokker–Planck–Kolmogorov equation (see also [11] and more generally [10]): writing $\mu_t := \mathscr{L}(x_t)$ and $a := \frac{1}{2}\sigma\sigma^T$, for $t \in I$

$$\langle \mu_t, \varphi \rangle = \langle \mu_0, \varphi \rangle + \int_0^t \left\langle \mu_s, b(s, \cdot, \mu_s)\partial_x \varphi + \mathrm{tr}\left(a(s, \cdot, \mu_s)\partial_x^2 \varphi \right) \right\rangle\, ds \quad \forall \varphi \in C_0^2(D)\,. \tag{2.1.2}$$

The aim of this chapter is to study the existence and uniqueness of solutions to the equation (2.1.1). We will show that a weak solution to (2.1.1) exists for unbounded and continuous coefficients, provided that we can find an appropriate measure-dependent Lyapunov function. The work on SDEs with coefficients that depend on the law of the solution was initiated by McKean [80]. In [93], Sznitman showed that if the coefficients of (2.1.1) are globally Lipschitz continuous, a fixed point argument on Wasserstein space can be carried out, and consequently a solution to (2.1.1) is obtained as the limit of classical SDEs. To extend this result, Funaki [39] formulated a non-linear martingale problem for McKean–Vlasov SDEs that allowed him to establish the existence of a solution to (2.1.1) by studying a limiting law of Euler approximations. His proof of existence holds for continuous coefficients satisfying a Lyapunov type condition in the state variable $x \in \mathbb{R}^d$ with polynomial Lyapunov functions. Whilst we also assume continuity of the coefficients, we allow for a much more general Lyapunov condition that depends on a measure. Furthermore, Funaki uses Lyapunov functions to establish integrability of the Euler scheme which is problematic if one wants to depart from polynomial functions, see [94]. Gärtner [40], uses an integrated Lyapunov condition with a Lyapunov function not dependent on measure, to study the weak well-posedness of McKean–Vlasov SDEs.

As explained in the previous chapter, an alternative approach to establishing existence of solutions to McKean–Vlasov equations is to approximate the equation with a particle system (a system of classical SDEs that interact with each other through empirical measure) and show that the limiting law solves the martingale problem. In this approach, one works with laws of empirical measures and proves their convergence to a (weak) solution of (2.1.1) by studying the corresponding non-linear martingale problem. We refer to [83] for a general overview and to [17, 38] and references within for recent results exploring this method. A general approach to establish the existence of martingale solutions has also been presented in [77]. We also refer the reader to interesting new developments on existence and uniqueness of solutions for McKean-Vlasov equations with non-smooth coefficients found in [84, 29]. Here, inspired by [84], we tackle the problem using the Skorokhod representation theorem and convergence lemma [91].

For classical SDEs (equations with no dependence on the law), the lack of sufficient regularity of the coefficients, say Lipschitz continuity, proves to be the main challenge in establishing existence and uniqueness of solutions. Lack of boundedness of the coefficients, typically, does not lead to significant difficulty, provided these are at least locally bounded. In this case, one can work with local solutions and the only concern is the possible explosion. Conditions that ensure that the solution does not explode may be formulated by using Lyapunov function techniques as has been pioneered in [58]. The key observation is that if one considers two SDEs with coefficients that agree on some bounded domain then the solutions, if unique, also agree until first time the solution leaves the domain, see, for example [92, Ch. 10].

This classical localisation procedure does not carry over, at least directly, from the setting of classical SDEs to McKean–Vlasov SDEs. Indeed, if we stop a classical SDE then until the stopping time the stopped process satisfies the same equation. If we take (2.1.1) and consider the stopped process $y_t := x_{t\wedge\tau}$, with some stopping time τ, then the equation this satisfies is

$$y_t = y_0 + \int_0^{t\wedge\tau} b(s, y_s, \mathscr{L}(x_s))\, ds + \int_0^{t\wedge\tau} \sigma(s, y_s, \mathscr{L}(x_s))\, dw_s\,, \quad t \in I\,.$$

Clearly, even for $t \le \tau$ this is not the same equation since $\mathscr{L}(x_s) \neq \mathscr{L}(y_s)$. This could be problematic if one would like to obtain a solution to McKean–Vlasov SDEs through a limiting procedure of stopped processes. Furthermore, let $D_k \subseteq D_{k+1}$ be a sequence of nested domains, and consider functions $\bar{b}$ and $\bar{\sigma}$ such that $\bar{b} = b$ and $\bar{\sigma} = \sigma$ on D_k. The equation

$$\bar{x}_t = \bar{x}_0 + \int_0^t \bar{b}(s, \bar{x}_s, \mathscr{L}(\bar{x}_s))\, ds + \int_0^t \bar{\sigma}(s, \bar{x}_s, \mathscr{L}(\bar{x}_s))\, dw_s\,, \quad t \in I\,,$$

is a McKean–Vlasov SDE, but in general $x_t \neq \bar{x}_t$ even for $t \le \bar{\tau}^k$, where $\bar{\tau}^k = \inf\{t \ge 0 : \bar{x}_t \notin D_k\}$. This implies that if one considers a sequence of SDEs with coefficients that agree on these subdomains, one no longer has monotonicity for the corresponding stopping times. We show that despite these difficulties it still possible to establish the existence of weak solutions to the McKean–Vlasov SDEs (2.1.1) using the idea of localisation, but extra care is needed.

2.1.1 Main Contributions

Our first main contribution is the generalisation of Lyapunov function techniques to the setting of McKean–Vlasov SDEs. The coefficients of the equation (2.1.1) depend on $(x, \mu) \in D \times \mathcal{P}(D)$ for $D \subseteq \mathbb{R}^d$. Hence the class of Lyapunov functions considered in this chapter also depend on $(x, \mu) \in D \times \mathcal{P}(D)$. See (2.2.1). Furthermore, it is natural to formulate the integrated Lyapunov condition, in which the key stability assumption is required to hold only on $\mathcal{P}(D)$, see (2.2.2) and Section 2.1.2 for motivating examples. Note that it is not immediately clear how one can obtain tightness estimates for the particle approximation under the integrated conditions we propose. To work with Lyapunov functions on $\mathcal{P}(D)$, we take advantage of the analysis on Wasserstein spaces, and in particular derivatives with respect to a measure as introduced by Lions in his lectures at Collège de France, see [22] and [26, Ch.5]. This analysis is presented in the appendix where the measure derivative in a domain is given. Furthermore, the calculus on Wasserstein spaces allows one to study a Fokker–Planck–Kolmogorov-type equation on $\mathcal{P}_2(D)$. Indeed, writing $\mu_t := \mathscr{L}(x_t)$ we have, for $\phi \in \mathcal{C}^{(1,1)}(\mathcal{P}_2(D))$ (see Definition A.3.1) and $t \in I$, that

$$\phi(\mu_t) = \phi(\mu_0) + \int_0^t \langle \mu_s, b(s,\cdot,\mu_s)\partial_\mu \phi(\mu_s) + \mathrm{tr}\left[a(s,\cdot,\mu_s)\partial_y \partial_\mu \phi(\mu_s)\right] \rangle \, ds. \tag{2.1.3}$$

Following the remark by Lions from his lectures at Collège de France, the equation (2.1.3) can be interpreted as a non-local transport equation on the space of measures. The reader may consult [26, Ch.5 Sec.7.4] for further details.

We formulate uniqueness results under the Lyapunov type condition and the integrated Lyapunov type condition that is required to hold only on $\mathcal{P}(D)$. This extends the standard monotone type conditions studied in literature e.g [14, 101, 79, 89]. Interestingly, in some special cases we are able to obtain uniqueness only under local monotone conditions. We support our results with the example inspired by Scheutzow [90] who has shown that, in general, uniqueness of solution to McKean–Vlasov SDEs does not hold if the coefficients are only locally Lipschitz. Finally, the results obtained in this chapter imply existence of a stationary solution to (2.1.3) in the case where b and σ do not depend on time. Note that we do not require a non-degeneracy condition on the diffusion coefficient.

2.1.2 Motivating Examples

Let us now present some examples to motivate the choice of the Lyapunov conditions. Consider first the McKean–Vlasov stochastic differential equation

$$dx_t = -x_t \left[\int_{\mathbb{R}} y^4 \mathscr{L}(x_t)(dy) \right] dt + \frac{1}{\sqrt{2}} x_t \, dw_t \,, \quad x_0 \in L^4(\mathcal{F}_0, \mathbb{R}) \,. \tag{2.1.4}$$

The diffusion generator for (2.1.4) is

$$L(x,\mu)v(x) := \frac{1}{4} x^2 v''(x) - x \left[\int_{\mathbb{R}} y^4 \, \mu(dy) \right] v'(x) \,. \tag{2.1.5}$$

It is not clear whether one can find a Lyapunov function such that the classical Lyapunov condition holds i.e. $L(x,\mu)v(x) \leq m_1 v(x) + m_2$, with $m_1 < 0$ in particular and $m_2 \in \mathbb{R}$. However, with the Lyapunov function given by $v(x) = x^4$ we can establish that

$$\int_{\mathbb{R}} L(x,\mu)v(x)\mu(dx) \leq -\int_{\mathbb{R}} v(x)\mu(dx) + 1. \tag{2.1.6}$$

See Example 2.2.14 for details. We will see that this is sufficient to establish integrability of (2.1.4) on $I = [0,\infty)$. See Theorem 2.2.10 and condition (2.2.7).

Another way to proceed, is to directly work with $v(\mu) := \int_{\mathbb{R}} x^4\,\mu(dx)$ as a Lyapunov function on the measure space $\mathcal{P}(\mathbb{R})$. This requires the use of derivatives with respect to a measure as introduced by Lions in his lectures at College de France, see [22] or Appendix A. We note that derivatives with respect to a measure are defined in $\mathcal{P}_2(\mathbb{R})$, and therefore one cannot apply Itô formula for arbitrary measures in $\mathcal{P}(D)$. However, we will only apply the Itô formula for measures supported on compact subsets of $\mathbb{R}^d$ and hence, measures that are in $\mathcal{P}_2(\mathbb{R})$. Then

$$\partial_\mu v(\mu)(y) = 4y^3, \quad \partial_y\partial_\mu v(\mu)(y) = 12y^2, \;\; y \in \mathbb{R}\,.$$

The generator corresponding to the appropriate Itô formula, see e.g. Proposition A.3.2, is

$$\begin{aligned} L^\mu v(\mu) &:= \int_{\mathbb{R}} \left(-x \int_{\mathbb{R}} y^4\,\mu(dy)\partial_\mu v(\mu)(x) + \frac{1}{4}x^2\partial_y\partial_\mu v(\mu)(x) \right) \mu(dx) \\ &= \int_{\mathbb{R}} \left(-4x^4 \int_{\mathbb{R}} y^4\mu(dy) + 3x^4 \right) \mu(dx)\,. \end{aligned}$$

We note that this yields the same expression as found when $v(x) = x^4$ in (2.1.5) after we integrate over μ (and so (2.1.6) again holds). In this case using the Itô formula for measure derivatives brings no advantages. However, the advantage of working with a Lyapunov function on the measure space appears when the dependence on the measure in the Lyapunov function is not linear.

Consider the following McKean–Vlasov stochastic differential equation

$$dx_t = -\left(\int_{\mathbb{R}} (x_t - \alpha y)\mathscr{L}(x_t)(dy) \right)^3 dt + \left(\int_{\mathbb{R}} (x_t - \alpha y)\mathscr{L}(x_t)(dy) \right)^2 \sigma\, dw_t\,, \tag{2.1.7}$$

for $t \in I$, α and σ constants and with $x_0 \in L^4(\mathcal{F}_0, \mathbb{R})$. Assume that $m := -(6\sigma^2 - 4 + 4\alpha) > 0$. Since the drift and diffusion are non-linear functions of the law and state of the process, it is natural to seek a Lyapunov function $v \in \mathcal{C}^{2,(1,1)}(\mathbb{R} \times \mathcal{P}(\mathbb{R}))$. See Definition A.3.3. The generator corresponding to the appropriate Itô formula, see e.g. Proposition A.3.4, is then given by (2.2.1) and we will show that for the Lyapunov function

$$v(x,\mu) = \left(\int_{\mathbb{R}} (x - \alpha y)\mu(dy) \right)^4\,,$$

we have

$$\int_{\mathbb{R}} (L^\mu v)(x,\mu)\,\mu(dx) \le m - m\int_{\mathbb{R}} v(x,\mu)\,\mu(dx)\,.$$

See Example 2.2.15 for details. Thus the condition (2.2.7) holds. This is sufficient to establish existence of solutions to (2.1.4) on $I = [0,\infty)$ as Theorem 2.2.10 will tell us.

Regarding our continuity assumptions for existence of solutions to (2.1.1) we note that we only require a type of joint continuity of the coefficients in $(x,\mu) \in D \times \mathcal{P}(D)$, and that allows us to consider coefficients where the dependence on the measure does not arise via the integration of an interaction kernel. This could be for example,

$$S_\alpha(\mu) := \frac{1}{\alpha}\int_0^\alpha \inf\{x \in \mathbb{R} \,:\, \mu((-\infty, x]) \ge s]\}\, ds\,,$$

for $\alpha > 0$ fixed. This quantity is known as the "expected shortfall" and is a type of risk measure. See Example 2.2.16 for details. These motivating examples also satisfy the Lyapunov type estimates, appearing in Section 2.3, that ensure the uniqueness of solutions.

2.2 Existence Results

For an open domain $D \subseteq \mathbb{R}^d$, we will use the notation $\mathcal{P}(D)$ for the space of probability measures over $(D, \mathcal{B}(D))$. We will consider this as a topological space with the topology of weak convergence. We will write $\mu_n \xrightarrow{w} \mu$ if $(\mu_n)_n$ converges to μ in the sense of weak convergence of probability measures. For $p \geq 1$ we use $\mathcal{P}_p(D)$ to denote the set of probability measures on D with finite p^{th} moment (i.e. $\int_D |x|^p \mu(dx) < \infty$ for $\mu \in \mathcal{P}_p(D)$). We will consider this as a metric space with the metric given by the p^{th} Wasserstein distance, see (2.2.8). Denote by $C_b(D)$ and $C_0(D)$ the subspaces of continuous functions that are bounded and compactly supported, respectively.

Recall that we are using the concept of derivatives with respect to a measure as introduced by Lions in his lectures at Collège de France, see [22]. For convenience, the construction and main definitions are in Appendix A. In particular, see Definition A.3.3 to clarify what is meant by the space $\mathcal{C}^{1,2,(1,1)}(I \times D \times \mathcal{P}(D))$. In short, saying that a function v is in this space means that all the derivatives appearing imminently in (2.2.1) exist and are appropriately jointly continuous so that we may apply the Itô formula for a function of a process and a flow of measures, see Proposition A.3.4. The use of such an Itô formula naturally leads to the following form of a diffusion generator. First, we note that throughout this chapter we assume that for the domain $D \subseteq \mathbb{R}^d$ there is a nested sequence of bounded sub-domains. By this we mean a sequence of bounded open subsets of $\mathbb{R}^d$, $(D_k)_k$ such that $\bigcup_k D_k = D$ and $\overline{D}_k \subset D_{k+1}$ for all k, i.e. $d(D_k, \partial D_{k+1}) := \inf_{x \in D_k, y \in \partial D_{k+1}} |x - y| > 0$ for all $k \in \mathbb{N}$. For $(t, x) \in I \times D$, $\mu \in \mathcal{P}(D_k)$ for some $k \in \mathbb{N}$ and for some $v \in \mathcal{C}^{1,2,(1,1)}(I \times D \times \mathcal{P}_2(D))$ we define the diffusion generator $L^\mu = L^\mu(t, x, \mu)$ as

$$
\begin{aligned}
&(L^\mu v)(t, x, \mu) \\
&:= \left(\partial_t v + \frac{1}{2}\mathrm{tr}\left(\sigma\sigma^T \partial_x^2 v\right) + b\partial_x v\right)(t, x, \mu) \\
&+ \int_{\mathbb{R}^d} \left(b(t, y, \mu)(\partial_\mu v)(t, x, \mu)(y) + \frac{1}{2}\mathrm{tr}\left(a(t, y, \mu)(\partial_y \partial_\mu v)(t, x, \mu)(y)\right)\right) \mu(dy).
\end{aligned}
\tag{2.2.1}
$$

We note that in the case $v \in C^{1,2}(I \times D)$, i.e when v does not depend on the measure, the above generator reduces to

$$
(L^\mu v)(t, x) = (Lv)(t, x) := \left(\partial_t v + \frac{1}{2}\mathrm{tr}\left(\sigma\sigma^T \partial_x^2 v\right) + b\partial_x v\right)(t, x).
$$

2.2.1 Assumptions and Main Result

We assume that $b : \mathcal{I} \times D \times \mathcal{P}(D) \to \mathbb{R}^d$ and $\sigma : \mathcal{I} \times D \times \mathcal{P}(D) \to \mathbb{R}^d \times \mathbb{R}^{d'}$ are measurable (later we will add joint continuity and local boundedness assumptions). We require the existence of a Lyapunov function satisfying one of the following conditions:

Assumption 2.2.1 (Lyapunov Condition)**.** There is a function $v \in \mathcal{C}^{1,2,(1,1)}(\mathcal{I} \times D \times \mathcal{P}_2(D))$, $v \geq 0$, and locally integrable, non-random functions $m_1 = m_1(t)$ and $m_2 = m_2(t)$ on $\mathcal{I}$ such that for any $k \in \mathbb{N}$, for all $t \in \mathcal{I}$, $x \in D_k$ and $\mu \in \mathcal{P}(D_k)$, we have,

$$L^{\mu}(t,x,\mu)v(t,x,\mu) \leq m_1(t)v(t,x,\mu) + m_2(t). \tag{2.2.2}$$

2.1a) We say that Lyapunov condition 2.2.1a holds if (2.2.2) holds and there is a non-negative function $V = V(t,x)$ such that for any $k \in \mathbb{N}$, for all $t \in \mathcal{I}$, $x \in D_k$ and all $\mu \in \mathcal{P}(D_k)$, we have,

$$V(t,x) \leq v(t,x,\mu) \tag{2.2.3}$$

and

$$V_k := \inf_{s \in \mathcal{I}, x \in \partial D_k} V(s,x) \to \infty \text{ as } k \to \infty. \tag{2.2.4}$$

2.1b) We say that Lyapunov condition 2.2.1b holds if (2.2.2) holds and there exists a non-negative function V such that for any $k \in \mathbb{N}$, for all $t \in \mathcal{I}$ and $\mu \in \mathcal{P}(D_k)$, we have,

$$\int_{D_k} V(t,x)\,\mu(dx) \leq \int_{D_k} v(t,x,\mu)\,\mu(dx) \tag{2.2.5}$$

and

$$V_k^c := \inf_{s \in \mathcal{I}, x \in D_k^c} V(s,x) \to \infty \text{ as } k \to \infty. \tag{2.2.6}$$

Assumption 2.2.2 (Integrated Lyapunov Condition)**.**
There is a $v \in \mathcal{C}^{1,2,(1,1)}(\mathcal{I} \times D \times \mathcal{P}_2(D))$, $v \geq 0$, such that:
i) There are locally integrable, non-random, functions $m_1 = m_1(t)$ and $m_2 = m_2(t)$ on $\mathcal{I}$ such that for any $k \in \mathbb{N}$, for all $t \in \mathcal{I}$ and $\mu \in \mathcal{P}(D_k)$, we have,

$$\int_{D_k} L^{\mu}(t,x,\mu)v(t,x,\mu)\mu(dx) \leq m_1(t)\int_{D_k} v(t,x,\mu)\mu(dx) + m_2(t) \tag{2.2.7}$$

ii) There is a non negative function $V = V(t,x)$ satisfying (2.2.5) and (2.2.6).

Assumption 2.2.3 (Initial Distribution)**.** We assume that for a given Lyapunov function v, the initial distribution $\mu_0 := \mathscr{L}(x_0)$ is such that μ_0 can be approximated by a sequence of probability distributions $(\mu_0^k)_k$ such that $\mu_0^k \xrightarrow{w} \mu_0$ and for each $k \in \mathbb{N}$, μ_0^k is supported on D_k and for some increasing continuous function $\varphi_v : [0,\infty) \to [0,\infty)$ such that $\varphi_v(x) \geq x$ for all $x \in [0,\infty)$ we have,

$$\langle \mu_0^k, v(0,\cdot,\mu_0^k)\rangle \leq \varphi_v(\langle \mu_0, v(0,\cdot,\mu_0)\rangle) < \infty.$$

Remark 2.2.4.

i) We have deliberately not specified the signs of the functions m_1 and m_2.

ii) Note that if μ_0 is supported on some D_K for any $K \in \mathbb{N}$, then Assumption 2.2.3 is satisfied after relabelling the sequence (D_k) to start from D_K and setting $\mu_0^k = \mu_0$.

iii) Regarding Assumption 2.2.3, it would be preferable to be able to prescribe an approximating sequence μ_0^k. It is easy to imagine however, how this condition should look in the case where $D = \mathbb{R}$ and $v(x,\mu) := x^2$. One simply truncates the measure μ_0 on $D_k := (-k,k)$ and puts the mass of the measure μ_0 outside D_k at the origin i.e. $\mu_0^k(dx) := \mathbb{1}_{x\in D_k}\mu_0(dx) + \mu_0(D_k^c)\delta_0$. The increasing continuous function φ_v in Assumption 2.2.3 facilitates the finding of such a Lyapunov function. The fact that searching for a Lyapunov function for a McKean-Vlasov SDE should also depend upon the initial distribution and not just the form of the coefficients should not be too surprising given the dependence of the coefficients on the law of the solution. Also, in [13], Example 1.1 shows that the existence and convergence to a stationary distribution of the non-linear Fokker Planck equation depends not only on the form of the measure dependence of the coefficients, but also on the initial condition.

Regarding the continuity of coefficients in (2.1.1) and their local boundedness we require the following.

Assumption 2.2.5 (v-Continuity). Functions $b : I \times D \times \mathcal{P}(D) \to \mathbb{R}^d$ and $\sigma : I \times D \times \mathcal{P}(D) \to \mathbb{R}^d \times \mathbb{R}^{d'}$ are jointly continuous in the last two arguments in the following sense: if $(\mu_n)_n \subset \mathcal{P}(D)$ are such that

$$\sup_n \sup_{t\in I} \int_D v(t,x,\mu_n)\,\mu_n(dx) < \infty$$

and if $(x_n \to x, \mu_n \overset{w}{\to} \mu)$ as $n \to \infty$ then for any $t \in I$, $b(t,x_n,\mu_n) \to b(t,x,\mu)$ and $\sigma(t,x_n,\mu_n) \to \sigma(t,x,\mu)$ as $n \to \infty$.

Assumption 2.2.6 (Local v-Boundedness). There exist constants $c_k \geq 0$ such that for any $\mu \in \mathcal{P}(D)$

$$\sup_{x\in D_k} |b(t,x,\mu)| \leq c_k \left(1 + \int_D v(t,y,\mu)\mu(dy)\right),$$

$$\sup_{x\in D_k} |\sigma(t,x,\mu)| \leq c_k \left(1 + \int_D v(t,y,\mu)\mu(dy)\right).$$

Assumption 2.2.7 (Integrated v-Growth). There exists an increasing function φ_c from $[0,\infty)$ to $[0,\infty)$ such that for all $\mu \in \mathcal{P}(D)$, we have,

$$\int_D |b(t,x,\mu)| + |\sigma(t,x,\mu)|^2 \mu(dx) \leq \varphi_c\left(\int_D v(t,x,\mu)\mu(dx)\right), \quad \forall t \in I.$$

Assumption 2.2.5 of v-continuity in the measure argument is very weak, but may in practice be hard to verify. In the case of unbounded domains, the property (2.2.5) will often hold for functions of the form $V(x) = |x|^p$, $p \geq 1$. In this situation, we have $\mu_n \in \mathcal{P}_p(D)$ for all the measures μ_n under consideration for convergence of the coefficients with a uniform bound on their p^{th} moments. However, from [100, Theorem 6.9], we know that for $(\mu_n)_n \subset \mathcal{P}_p(D)$ with uniform bound on their p^{th} moments, weak convergence of measures is equivalent to convergence in the p^{th} Wasserstein distance. Hence, in such case, it is enough to check that if $x_n \to x$ and $W_p(\mu_n, \mu) \to 0$ as $n \to \infty$ then $b(x_n, \mu_n) \to b(x, \mu)$ and $\sigma(x_n, \mu_n) \to \sigma(x, \mu)$ as $n \to \infty$. This will be satisfied in particular if

$$|b(x_n, \mu_n) - b(x, \mu)| + |\sigma(x_n, \mu_n) - \sigma(x, \mu)| \leq \rho(|x - x_n|) + W_p(\mu_n, \mu),$$

for some function $\rho = \rho(x)$ such that $\rho(|x|) \to 0$ as $x \to 0$. We note that this is a common assumption, see e.g. [39]. At this point it may be worth noting that the p^{th}-Wasserstein distance on $\mathcal{P}_p(D)$ is

$$W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu,\nu)} \int_{D \times D} |x - y|^p \, \pi(dx, dy) \right)^{\frac{1}{p}}, \qquad (2.2.8)$$

where $\Pi(\mu, \nu)$ denotes the set of couplings between μ and ν i.e. all measures on $(D \times D, \mathcal{B}(D \times D))$ such that $\pi(B, D) = \mu(B)$ and $\pi(D, B) = \nu(B)$ for every $B \in \mathcal{B}(D)$.

Note that in the case of McKean–Vlasov SDEs it is often useful to think of the solution as a pair, consisting of the process x and its law i.e. $(x_t, \mathscr{L}(x_t))_{t \in \mathcal{I}}$. The coefficients of the McKean–Vlasov SDE depend on the law of the solution and the main focus of this chapter is on equations with unbounded coefficients, therefore a condition on integrability of the law is natural.

Definition 2.2.8 (v-Integrable Weak Solution).
A v-integrable weak solution to (2.1.1), on $\mathcal{I}$ in D is

$$(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in \mathcal{I}}, (w_t)_{t \in \mathcal{I}}, (x_t)_{t \in \mathcal{I}}),$$

where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, $(\mathcal{F}_t)_{t \in \mathcal{I}}$ is a filtration, $(w_t)_{t \in \mathcal{I}}$ is an $(\mathcal{F}_t)_{t \in \mathcal{I}}$-Wiener process, $(x_t)_{t \in \mathcal{I}}$ is an adapted process satisfying (2.1.1) such that $x \in C(\mathcal{I}; D)$ a.s. and finally, for all $t \in \mathcal{I}$ we have $\mathbb{E}[v(t, x_t, \mathscr{L}(x_t))] < \infty$.

Before we state the main theorem of this chapter, we state the conditions on m_1, m_2 that allow one to establish the requisite uniform estimates for our approximations, which, in the case where $\mathcal{I} = [0, \infty)$ need to be uniform in time.

Define $\gamma(t) := \exp\left(-\int_0^t m_1(s)\, ds\right)$ and

$$\begin{aligned} M(t) :=& \frac{\varphi_v(\langle \mu_0, v(0, \cdot, \mu_0) \rangle)}{\gamma(t)} + \int_0^t \frac{\gamma(s)}{\gamma(t)} m_2(s) ds, \\ M^+(t) :=& e^{\int_0^t (m_1(s))^+ \, ds} \left(\varphi_v(\langle \mu_0, v(0, \cdot, \mu_0) \rangle) + \int_0^t \gamma(s) m_2^+(s) \, ds \right). \end{aligned} \qquad (2.2.9)$$

Note that $M(t) \leq M^+(t)$.

Remark 2.2.9 (Conditions on m_1 and m_2 Ensuring Finiteness of M^+).

i) If $I=[0,T]$, m_1 and m_2 are set to 0 outside I, leading to

$$\sup_{t<\infty}\int_0^t \frac{\gamma(s)}{\gamma(t)} m_2(s)ds \le \int_0^T e^{\int_s^T m_1(r)\,dr}|m_2(s)|ds < \infty\,.$$

ii) If $I=[0,\infty)$,

$$m_1(t)\le 0 \;\; \forall t\ge 0 \;\text{ and }\; \int_0^\infty |m_2(s)|\,ds < \infty\,, \tag{2.2.10}$$

then

$$\sup_{t<\infty}\int_0^t e^{\int_s^t m_1(r)\,dr} m_2(s)\,ds \le \int_0^\infty |m_2(s)|\,ds < \infty.$$

In both of these cases we have $\sup_{t\in I} M(t)<\infty$ and $\sup_{t\in I} M^+(t)<\infty$.

Theorem 2.2.10. *Let $D\subseteq\mathbb{R}^d$ and Assumptions 2.2.3, 2.2.5 and 2.2.6 hold. Then the following statements are true:*

i) If Assumption 2.2.1a holds and $\sup_{t\in I} M^+(t)<\infty$, then there exists a v-integrable weak solution to (2.1.1) on I.

ii) Let either Assumption 2.2.1b or Assumption 2.2.2 hold. If additionally Assumption 2.2.7 holds and $\sup_{t\in I} M(t)<\infty$, then there exists a v-integrable weak solution to (2.1.1) on I.

In all of the above cases we also have,

$$\sup_{t\in I}\mathbb{E}[v(t,x_t,\mathscr{L}(x_t))]<\infty\,.$$

We make the following comment. By virtue of Assumption 2.2.6 we have that under the conditions of Theorem 2.2.10, the v-integrable weak solution to (2.1.1) obtained by the theorem satisfies the forward nonlinear Fokker–Planck–Kolmogorov equation (2.1.2), where $\mu_t=\mathscr{L}(x_t)$.

2.2.2 Proof of the Existence Results

We will use the convention that the infimum of an empty set is positive infinity. We extend b and σ in a measurable but discontinuous way to functions on $\mathbb{R}^+\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)$ by taking

$$b(t,x,\mu)=\sigma(t,x,\mu)=0 \text{ if } x\in\mathbb{R}^d\setminus D \text{ or if } t\notin I.$$

For $t\notin I$ we set $m_1(t)=m_2(t)=0$. Consequently from here onwards let $I=[0,\infty)$. We define

$$b^k(t,x,\mu):=\mathbb{1}_{x\in D_k} b(t,x,\mu) \text{ and } \sigma^k(t,x,\mu):=\mathbb{1}_{x\in D_k}\sigma(t,x,\mu)\,.$$

We now provide some results (Lemmas 2.2.11 and 2.2.13 and Corollary 2.2.12) regarding a sequence of processes whose existence will be proved as part of Theorem 2.2.10.

Lemma 2.2.11. *Let Assumptions 2.2.3 and 2.2.6 hold. Let there exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $(\mathcal{F}_t)_{t\in I}$, an adapted Wiener process w and adapted processes $(x^k)_k$ that satisfy, for all $t \in I$,*

$$dx_t^k = b^k(t, x_t^k, \mathscr{L}(x_t^k))\, dt + \sigma^k(t, x_t^k, \mathscr{L}(x_t^k))\, dw_t\,, \quad \mathscr{L}(x_0^k) = \mu_0^k\,. \tag{2.2.11}$$

For any $m, k \in \mathbb{N}$, let $\tau_m^k := \inf\{t \in I : x_t^k \notin D_m\}$.

i) If either Assumption 2.2.1a, Assumption 2.2.1b or 2.2.2 hold then for any $t \in I$,

$$\sup_k \mathbb{E}[v(t, x_t^k, \mathscr{L}(x_t^k))] \le M(t)\,.$$

ii) If either Assumption 2.2.1a, Assumption 2.2.1b or 2.2.2 hold then for any $t \in I$ and $k \in \mathbb{N}$,

$$\mathbb{P}(\tau_k^k < t) \le M(t)V_k^{-1}\,.$$

iii) If Assumption 2.2.1a holds then for any $t \in I$,

$$\sup_k \mathbb{P}(\tau_m^k < t) \le M^+(t)V_m^{-1} + \mathbb{P}(x_0^k \in D_m).$$

iv) If Assumption 2.2.1b or Assumption 2.2.2 holds then for any $t \in I$,

$$\sup_k \mathbb{P}(x_t^k \notin D_m) \le M(t)V_m^{-1}\,.$$

Proof. i) Since μ_0^k is supported on D_k for each $k \in \mathbb{N}$ and the coefficients b^k and σ^k are zero outside D_k, we know that the support of x_t^k is contained within $\overline{D}_k$ for all $t \in I$. Therefore $\mathscr{L}(x_t^k) \in \mathcal{P}_2(D)$, and we can apply the Itô formula from Proposition A.3.4 to γv with arguments t, x^k and its law. Thus

$$\begin{aligned}\gamma(t)v(t, x_t^k, \mathscr{L}(x_t^k)) &= \gamma(0)v(0, x_0^k, \mathscr{L}(x_0^k)) \\ &\quad + \int_0^t \gamma(s)[L^\mu v - m_1 v](s, x_s^k, \mathscr{L}(x_s^k))\, ds + \int_0^t \gamma(s)[(\partial_x v)\sigma](s, x_s^k, \mathscr{L}(x_s^k))\, dw_s\,.\end{aligned}$$

Due to the local boundedness of the coefficients and either Lyapunov condition (2.2.2) or (2.2.7) combined with Remark 2.2.4 ii) we get

$$\mathbb{E}[\gamma(t)v(t, x_t^k, \mathscr{L}(x_t^k))] \le \mathbb{E}[\gamma(0)v(0, x_0^k, \mathscr{L}(x_0^k))] + \int_0^t \gamma(s)m_2(s)ds\,. \tag{2.2.12}$$

This proves the first part of the lemma.

ii) For the second part, noting that the coefficients b^k and σ^k are zero outside D_k, once the process x^k leaves D_k the process stops, yielding $x_t^k = x_{t\wedge\tau_k^k}^k$ for all $t \in I$, which implies $\mathscr{L}(x_t^k) = \mathscr{L}(x_{t\wedge\tau_k^k}^k)$ for all $t \in I$. We further observe using (2.2.5) that,

$$\begin{aligned}\mathbb{E}[v(t, x_t^k, \mathscr{L}(x_t^k))] = \mathbb{E}[v(t, x_{t\wedge\tau_k^k}^k, \mathscr{L}(x_{t\wedge\tau_k^k}^k))] &\ge \mathbb{E}[V(t, x_{t\wedge\tau_k^k}^k)\mathbb{1}_{\tau_k^k < t}] \\ &= \mathbb{E}[V(t, x_{\tau_k^k}^k)\mathbb{1}_{\tau_k^k < t}] \\ &\ge V_k \mathbb{P}(\tau_k^k < t)\,.\end{aligned}$$

Hence,

$$\mathbb{P}(\tau_k^k < t) \leq \frac{\mathbb{E}[v(t, x^k_{t\wedge\tau_k^k}, \mathscr{L}(x_t^k))]}{V_k}\,.$$

This completes the proof of the second statement.

iii) To prove the third statement we first note that for $m > k$ we have $\mathbb{P}(\tau_m^k < t) = \mathbb{P}(x_0^k \notin D_m) = 0$. Thus we may assume that $m \leq k$. We proceed similarly as above but with the crucial difference that x_t^k is no longer equal to $x^k_{t\wedge\tau_k^k}$. Our aim is to apply the Itô formula to the function v, evaluated over the process $(x^k_{t\wedge\tau_m^k})_{t\in I}$ and the flow of marginal measures $(\mathscr{L}(x_t^k))_{t\in I}$. Note that $\mathscr{L}(x^k_{t\wedge\tau_m^k}) \neq \mathscr{L}(x_t^k)$. Nevertheless the Itô formula A.3.4 may be applied. After taking expectations this yields

$$\begin{aligned}&\mathbb{E}\big[\gamma(t\wedge\tau_m^k)v(t\wedge\tau_m^k, x^k_{t\wedge\tau_m^k}, \mathscr{L}(x_t^k))\big]\\ &\quad=\mathbb{E}[v(0, x_0^k, \mathscr{L}(x_0^k))] + \mathbb{E}\bigg[\int_0^{t\wedge\tau_m^k} \gamma(s)\big[L^{\mu,k}v - m_1 v\big](s, x_s^k, \mathscr{L}(x_s^k))\,ds\bigg].\end{aligned}$$

We now use (2.2.2) to see that

$$\begin{aligned}&\mathbb{E}\left[\gamma(t\wedge\tau_m^k)v(t\wedge\tau_m^k, x^k_{t\wedge\tau_m^k}, \mathscr{L}(x_t^k))\right]\\ &\leq \mathbb{E}[v(0, x_0^k, \mathscr{L}(x_0^k))] + \mathbb{E}\bigg[\int_0^{t\wedge\tau_m^k}\gamma(s)m_2(s)\,ds\bigg]\\ &\leq \varphi_v(\langle\mu_0, v(0,\cdot,\mu_0)\rangle) + \int_0^t \gamma(s)m_2^+(s)\,ds =: \bar{M}(t)\,.\end{aligned}$$

Then

$$\inf_{s\leq t}\gamma(s)\mathbb{E}[v(t\wedge\tau_m^k, x^k_{t\wedge\tau_m^k}, \mathscr{L}(x_t^k))] \leq \mathbb{E}[\gamma(t\wedge\tau_m^k)v(t\wedge\tau_m^k, x^k_{t\wedge\tau_m^k}, \mathscr{L}(x_t^k))] \leq \bar{M}(t)$$

and so using (2.2.3) we see the following,

$$\begin{aligned}\mathbb{E}[v(t\wedge\tau_m^k, x^k_{t\wedge\tau_m^k}, \mathscr{L}(x_t^k))] \geq \mathbb{E}[V(t\wedge\tau_m^k, x^k_{t\wedge\tau_m^k})] &\geq\mathbb{E}[V(t\wedge\tau_m^k, x^k_{t\wedge\tau_m^k})\mathbb{1}_{\{0<\tau_m^k<t\}}]\\ &\geq V_m\mathbb{P}(0<\tau_m^k<t).\end{aligned}$$

Combining the above we have,

$$\mathbb{P}(\tau_m^k < t) =\mathbb{P}(0<\tau_m^k<t) + \mathbb{P}(\tau_m^k = 0) \leq \frac{1}{\inf_{s\leq t}\gamma(s)}\frac{\bar{M}(t)}{V_m} + \mathbb{P}(x_0^k \notin D_m)\,.$$

We conclude by observing that

$$\inf_{s\leq t}\gamma(s) \geq e^{-\int_0^t (m_1(s))^+\,ds}.$$

iv) To prove the fourth statement, first note that for $m > k$, $\mathbb{P}(x_t^k \notin D_m) = 0$ and hence we take $m \leq k$. Conditions (2.2.5) and (2.2.6) imply that

$$\begin{aligned}\mathbb{E}[v(t, x_t^k, \mathscr{L}(x_t^k))] \geq \int_D V(t,x)\mathscr{L}(x_t^k)(dx) &\geq \int_{D\cap D_m^c} V(t,x)\mathscr{L}(x_t^k)(dx)\\ &\geq V_m\mathbb{P}(x_t^k\notin D_m)\,.\end{aligned}$$

□

Since we are assuming that $\mathbb{P}(x_0 \in D) = 1$ we have

$$\lim_{k\to\infty} \mathbb{P}(x_0 \notin D_k) = 1 - \lim_{k\to\infty} \mathbb{P}(x_0 \in D_k) = 1 - \mathbb{P}\Big(\bigcup_k \{x_0 \in D_k\}\Big) = 0\,.$$

Corollary 2.2.12. *Let Assumption 2.2.6 hold. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in I}, \mathbb{P})$ be a filtered probability space equipped with an $(\mathcal{F}_t)_{t\in I}$-Wiener process w and a sequence of adapted processes $(x^k)_k$ such that (2.2.11) holds for all $t \in I$, $k \in \mathbb{N}$. Assume that $x^k \to x$ in $C(I; D)$ $\mathbb{P}$-almost surely. If either Assumption 2.2.1a, 2.2.1b or 2.2.2 hold then*

$$\sup_{t\in I} \mathbb{E}[v(t, x_t, \mathscr{L}(x_t))] \le \sup_{t\in I} M(t)\,,$$

where M is given in (2.2.9).

Proof. By Fatou's lemma, continuity of v and (2.2.9) we get

$$\mathbb{E}[v(t, x_t, \mathscr{L}(x_t))] \le \liminf_{k\to\infty} \mathbb{E}[v(t, x_t^k, \mathscr{L}(x_t^k))] \le \sup_{t\in I} M(t)\,.$$

The result follows after taking supremum over t. □

Our aim is to use Skorokhod's method to prove the existence of a weak solution to the equation (2.1.1). Before we proceed to the proof of Theorem 2.2.10, we need to establish tightness of the family of laws of the processes defined by (2.2.11).

Lemma 2.2.13 (Tightness). *Let Assumptions 2.2.3 and 2.2.6 hold and let there exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $(\mathcal{F}_t)_{t\in I}$, adapted Wiener process w and adapted processes $(x^k)_k$ that satisfy, for all $t \in I$, (2.2.11).*

i) If Assumption 2.2.1a holds with $\sup_{t\in I} M^+(t) < \infty$, then the law of $(x^k)_k$ is tight on $C(I; D)$.

ii) Let Assumptions 2.2.1b or 2.2.2 hold, with Assumption 2.2.7 and $\sup_{t\in I} M(t) < \infty$, then the law of $(x^k)_k$ is tight on $C(I; D)$. Additionally for any $\varepsilon > 0$, there is m_ε such that for all $m \ge m_\varepsilon$

$$\sup_k \mathbb{P}(\tau_m^k \in I) \le \varepsilon.$$

Proof. i) Under the Assumption 2.2.1a tightness of the law of $(x^k)_k$ on $C(I; D)$ follows from the third statement in Lemma 2.2.11. Indeed given $\varepsilon > 0$ we can find m_0 such that for any $m > m_0$

$$\sup_{t\in I} \mathbb{P}(\limsup_{k\to\infty} \tau_m^k < t) \le \sup_{t\in I} \liminf_{k\to\infty} \mathbb{P}(\tau_m^k < t) \le \sup_{t\in I} M^+(t) V_m^{-1} + \liminf_{k\to\infty} \mathbb{P}(x_0^k \notin D_m) \\ \le \varepsilon,$$

due to, in particular, our assumption that $V_m \to \infty$ as $m \to \infty$. By considering Assumption 2.2.6 of local v-boundedness along with the first statement of Lemma

2.2.11, the stopped processes $x^k_{\cdot\wedge\tau^k_m}$ are tight in $C([0,t];D)$ for any $t \in I$. Thus, as $\sup_{t\in I}\mathbb{P}(\limsup_{k\to\infty}\tau^k_m < t) \to 0$ as $m\to\infty$, we recover tightness of x^k in $C(I;D)$.

ii) First we observe that for every ℓ and $(t_1,\ldots,t_\ell)$ in I, the joint distribution of $(x^k_{t_1},\ldots,x^k_{t_\ell})$ is tight. Indeed, statement iv) in Lemma 2.2.11 guarantees tightness of the law of x^k_t for any $t\in I$. Given $\varepsilon>0$, for any $\ell\in\mathbb{N}$ we can find m_0 such that for any $m>m_0$

$$\mathbb{P}(x^k_{t_1}\notin D_m,\ldots,x^k_{t_\ell}\notin D_m) \le \ell\sup_{t\in I}M(t)V_m^{-1}\le\varepsilon\,,$$

due to the assumption that $V_m\to\infty$ as $m\to\infty$. We will use Skorokhod's Theorem (see [91, Ch. 1 Sec. 6]). This will allow us to conclude tightness of the law of $(x^k)_k$ on $C(I;D)$ as long as we can show that for any $\varepsilon>0$

$$\lim_{h\to0}\sup_k\sup_{|s_1-s_2|\le h}\mathbb{P}(|x^k_{s_1}-x^k_{s_2}|>\varepsilon)=0\,.$$

From (2.2.11), using the Assumption 2.2.7, we get, for $0<|s_1-s_2|<1$,

$$\begin{aligned}
&\mathbb{E}[|x^k_{s_1}-x^k_{s_2}|]\\
&\le\int_{s_2}^{s_1}\mathbb{E}[|b^k(r,x^k_r,\mathscr{L}(x^k_r))|]\,dr+\left(\mathbb{E}\left[\int_{s_2}^{s_1}|\sigma^k(r,x^k_r,\mathscr{L}(x^k_r))|^2\,dr\right]\right)^{\frac12}\\
&\le c\int_{s_2}^{s_1}\varphi_c(\sup_k\mathbb{E}[v(r,x^k_r,\mathscr{L}(x^k_r))])\,dr+\left(c\int_{s_2}^{s_1}\varphi_c(\sup_k\mathbb{E}[v(r,x^k_r,\mathscr{L}(x^k_r))])\,dr\right)^{\frac12}\\
&\le c\left(1+\varphi_c(\sup_{t\in I}M(t))\right)|s_1-s_2|^{\frac12}\,.
\end{aligned}$$

Markov's inequality leads to

$$\sup_k\sup_{|s_1-s_2|\le h}\mathbb{P}\left(|x^k_{s_1}-x^k_{s_2}|>\varepsilon\right)\le c\varepsilon h^{\frac12}$$

which concludes the proof of tightness.

We will now prove the second statement in ii). Note that $C(I;D)$ is open and $C(I;D_{k-1})\subset C(I;D_k)$ and $\bigcup_k C(I;D_k)=C(I;D)$. We know that for any $\varepsilon>0$ there is a compact set $\mathcal{K}_\varepsilon\subset C(I;D)$ such that

$$\sup_k\mathbb{P}(x^k\notin\mathcal{K}_\epsilon)\le\epsilon.$$

Since $\mathcal{K}_\varepsilon\subset C(I;D)$ is compact and the set of $(C(I;D_k))_k$ is an open cover, there must be some m_ε such that $\mathcal{K}_\varepsilon\subset C(I;D_{m_\varepsilon})$. But this means that

$$\mathbb{P}(x^k\notin C(I;D_{m_\varepsilon}))\le\mathbb{P}(x^k\notin\mathcal{K}_\varepsilon)$$

and so $\mathbb{P}(\tau^k_m\in I)=\mathbb{P}(x^k\notin C(I;D_m))\le\mathbb{P}(x^k\notin C(I;D_{m_\varepsilon}))\le\mathbb{P}(x^k\notin\mathcal{K}_\varepsilon)\le\varepsilon$ for all $m\ge m_\varepsilon$. □

Proof of Theorem 2.2.10. Let us define $t_i^n := \frac{i}{n}$, $i = 0, 1, \ldots$ and $\kappa_n(t) = t_i^n$ for $t \in [t_i^n, t_{i+1}^n)$. Fix k. We introduce Euler approximations $x^{k,n}$, $n \in \mathbb{N}$,

$$x_t^{k,n} = x_0 + \int_0^t b^k\left(s, x_{\kappa_n(s)}^{k,n}, \mathscr{L}(x_{\kappa_n(s)}^{k,n})\right) ds + \int_0^t \sigma^k\left(s, x_{\kappa_n(s)}^{k,n}, \mathscr{L}(x_{\kappa_n(s)}^{k,n})\right) dw_s \,.$$

Let us outline the proof: As a first step we fix k and we show tightness of the family of laws of the Euler approximations and apply Skorokhod's theorem to let $n \to \infty$ in the above equation, obtaining solutions to the truncated SDE. The second step is then to use Lemma 2.2.13 to show tightness with respect to k. Finally we can use Skorokhod's theorem again to show that (for a subsequence) the limit as $k \to \infty$ satisfies (2.1.1) (on a new probability space).

First Step. Using standard arguments, we can verify that, for a fixed k, the sequence $(x^{k,n})_n$ is tight (in the sense that the laws induced on $C([0,\infty), D)$ are tight). By Prohorov's theorem (see e.g. [8, Ch. 1, Sec. 5]), there is a subsequence (which we do not distinguish in notation) such that $\mathscr{L}(x^{k,n}) \xrightarrow{w} \mathscr{L}(x^k)$ as $n \to \infty$.

Hence we may apply Skorokhod's Representation Theorem (see e.g. [8, Ch. 1, Sec. 6]) and obtain a new probability space $(\tilde{\Omega}^k, \tilde{\mathcal{F}}^k, \tilde{\mathbb{P}}^k)$ where on this space there are new random elements $(\tilde{x}_0^n, \tilde{x}^{k,n}, \tilde{w}^n)$ and $(\tilde{x}_0, \tilde{x}^k, \tilde{w})$ such that

$$\mathscr{L}(\tilde{x}_0^n, \tilde{x}^{k,n}, \tilde{w}^n) = \mathscr{L}(x_0, x^{k,n}, w) \quad \forall n \in \mathbb{N}, \qquad \mathscr{L}(\tilde{x}_0, \tilde{x}^k, \tilde{w}) = \mathscr{L}(x_0, x^k, w) \quad \text{and}$$

$$(\tilde{x}_0^n, \tilde{x}^{k,n}, \tilde{w}^n) \to (\tilde{x}_0, \tilde{x}^k, \tilde{w}) \text{ as } n \to \infty \text{ in } C([0,\infty), D \times D \times \mathbb{R}^{d'}) \text{ surely.}$$

We let

$$\tilde{\mathcal{F}}_t^k := \sigma\{\tilde{x}_0\} \vee \sigma\{\tilde{x}_s, \tilde{w}_s : s \leq t\}$$

and define $\tilde{\mathcal{F}}_t^{k,n}$ analogously. Then $\tilde{w}^n$ and $\tilde{w}$ are $(\tilde{\mathcal{F}}_t^n)_{t\geq 0}$ and $(\tilde{\mathcal{F}}_t)_{t\geq 0}$-Wiener processes, respectively. Define

$$\tilde{\tau}_k^{k,n} := \inf\{t \geq 0 : \tilde{x}_t^{k,n} \notin D_k\} \text{ and } \tilde{\tau}_k^k := \inf\{t \geq 0 : \tilde{x}_t^k \notin D_k\} \,.$$

These are $\tilde{\mathcal{F}}^{k,n}$ and $\tilde{\mathcal{F}}^k$ stopping times respectively. Moreover, due to the convergence of the trajectories $\tilde{x}^{k,n}$ to $\tilde{x}^k$ we can see that,

$$\liminf_{n\to\infty} \tilde{\tau}_k^{k,n} \geq \tilde{\tau}_k^k \,.$$

From the fact that the laws of the sequences are identical we see that we still have the Euler approximation equation on the new probability space: for $t \geq 0$

$$d\tilde{x}_t^{k,n} = b^k(t, \tilde{x}_{\kappa_n(t)}^{k,n}, \mathscr{L}(\tilde{x}_{\kappa_n(t)}^{k,n}))\, dt + \sigma^k(t, \tilde{x}_{\kappa_n(t)}^{k,n}, \mathscr{L}(\tilde{x}_{\kappa_n(t)}^{k,n}))\, d\tilde{w}_t^n \,.$$

Using Skorohod's Lemma, see [91, Ch. 2, Sec. 3], together with the continuity conditions in Assumption 2.2.5, we can take $n \to \infty$ and conclude that for all $t \leq \tilde{\tau}_k^k$ we have

$$d\tilde{x}_t^k = b^k(t, \tilde{x}_t^k, \mathscr{L}(\tilde{x}_t^k))\, dt + \sigma^k(t, \tilde{x}_t^k, \mathscr{L}(\tilde{x}_t^k))\, d\tilde{w}_t \,. \tag{2.2.13}$$

At this point we remark that the process $\tilde{x}^k$ stopped at $\tilde{\tau}_k^k$, is well defined, continuous on $[0,\infty)$ and satisfies (2.2.13) for $t \in I$. Abusing notation, let $\tilde{x}^k$ refer to this stopped process. Additionally, it satisfies the equation without the cutting applied to the coefficients i.e for all $t \leq \tilde{\tau}^k_k$:

$$d\tilde{x}^k_t = b\left(t, \tilde{x}^k_t, \mathscr{L}(\tilde{x}^k_t)\right)\, dt + \sigma\left(t, \tilde{x}^k_t, \mathscr{L}(\tilde{x}^k_t)\right)\, d\tilde{w}_t\,.$$

Second Step. Tightness of the law of $(\tilde{x}^k)_k$ in $C(I;\bar{D})$ follows from Lemma 2.2.13 and Remark 2.2.9. From Prohorov's theorem there is a subsequence (again, not distinguished in notation) $\mathscr{L}(\tilde{x}^k) \xrightarrow{w} \mathscr{L}(\tilde{x})$ as $k \to \infty$. From Skorokhod's Representation Theorem we then obtain a new probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$ supporting random elements $(\bar{x}^k_0, \bar{x}^k, \bar{w}^k)$ and $(\bar{x}_0, \bar{x}, \bar{w})$ such that

$$\mathscr{L}(\bar{x}_0, \bar{x}, \bar{w}) = \mathscr{L}(\tilde{x}_0, \tilde{x}, \tilde{w})\,,$$

$$\mathscr{L}(\bar{x}^k_0, \bar{x}^k, \bar{w}^k) = \mathscr{L}(\tilde{x}^k_0, \tilde{x}^k, \tilde{w}^k) \quad \forall k \in \mathbb{N},$$

and

$$(\bar{x}^k_0, \bar{x}^k, \bar{w}^k) \to (\bar{x}_0, \bar{x}, \bar{w}) \text{ as } k \to \infty \text{ in } C(I; D \times \bar{D} \times \mathbb{R}^{d'}) \text{ surely}\,.$$

Let $\bar{\tau}^k_k := \inf\{t : \bar{x}^k_t \notin D_k\}$, $\bar{\tau}^k_m := \inf\{t : \bar{x}^k_t \notin D_m\}$ and $\bar{\tau}^\infty_m := \inf\{t : \bar{x}_t \notin D_m\}$. Since $\sup_{t<\infty} |\bar{x}^k_t - \bar{x}_t| \to 0$ we get $\limsup_{k\to\infty} \bar{\tau}^k_{m-1} \leq \bar{\tau}^\infty_m$ surely. To see why this holds, assume the contrary for finite $\bar{\tau}^\infty_m(\omega)$ since the infinite case holds immediately. We assume for a contradiction that $\limsup_{k\to\infty} \bar{\tau}^k_{m-1}(\omega) > \bar{\tau}^\infty_m(\omega)$. Then, there exists a subsequence k_j such that $\bar{\tau}^{k_j}_{m-1} > \bar{\tau}^\infty_m$ for all $j \in \mathbb{N}$. Consequently, $|\bar{x}^{k_j}_{\bar{\tau}^\infty_m} - \bar{x}_{\bar{\tau}^\infty_m}| \geq d(D_{m-1}, \partial D_m) > 0$. However, $|x^{k_j}_{\bar{\tau}^\infty_m} - \bar{x}_{\bar{\tau}^\infty_m}| \leq \sup_{t<\infty} |\bar{x}^{k_j}_t - \bar{x}_t| \to 0$ as $k_j \to \infty$ and we have arrived at a contradiction.

Then from Fatou's Lemma, and either part iii) of Lemma 2.2.11 or part ii) of Lemma 2.2.13 depending on the type of Lyapunov condition that holds, we have that, for any $s, t \in I$, $t < s$,

$$\begin{aligned}\mathbb{P}(\bar{\tau}^\infty_m \leq t) \leq \mathbb{P}((\limsup_{k\to\infty} \bar{\tau}^k_{m-1}) < s) &\leq \mathbb{P}(\liminf_{k\to\infty}\{\bar{\tau}^k_{m-1} < s\}) \\ &\leq \liminf_{k\to\infty} \mathbb{P}(\bar{\tau}^k_{m-1} < s) \\ &\leq \sup_k \mathbb{P}(\bar{\tau}^k_{m-1} \in I) \to 0 \text{ as } m \to \infty.\end{aligned} \tag{2.2.14}$$

Then the distribution of $\bar{\tau}^\infty_m$ converges in distribution, as $m \to \infty$, to a random variable $\bar{\tau}$ with distribution $\mathbb{P}(\bar{\tau} \leq T) = 0$ for any $T < \infty$ and $\mathbb{P}(\bar{\tau} = \infty) = 1$. In general, convergence in distribution does not imply convergence in probability. But in the special case that the limiting distribution corresponds to an almost surely constant random variable we obtain convergence in probability (see e.g. [36, Ch. 11, Sec. 1]). Hence $\bar{\tau}^\infty_m \to \infty$ in probability as $m \to \infty$. From this we can conclude that there is a subsequence that converges almost surely.

Since (2.2.13) holds for $\tilde{x}^k$ we have the corresponding equation for $\bar{x}^k$ i.e. for $t \leq \bar{\tau}^k_k$,

$$d\bar{x}^k_t = b(t, \bar{x}^k_t, \mathscr{L}(\bar{x}^k_t))\, dt + \sigma(t, \bar{x}^k_t, \mathscr{L}(\bar{x}^k_t))\, d\bar{w}^k_t\,. \tag{2.2.15}$$

Fix $m < k'$. We will consider $k > k'$. Then (2.2.15) holds for all $t \leq \inf_{k\geq k'} \bar{\tau}^k_m$. We can now consider $\bar{x}^k_{t\wedge\tau^k_m}$ (these all stay inside D_m for all $k > k' > m$) and use dominated convergence theorem for the bounded variation integral and Skorokhod's lemma on convergence of stochastic integrals, see [91, Ch. 2, Sec. 3], and our assumptions on continuity of b and σ to let $k \to \infty$. We obtain, for $t \le \inf_{k \ge k'} \bar{\tau}_m^k \wedge \bar{\tau}_m^\infty$,

$$d\bar{x}_t = b(t, \bar{x}_t, \mathscr{L}(\bar{x}_t))\, dt + \sigma(t, \bar{x}_t, \mathscr{L}(\bar{x}_t))\, d\bar{w}_t\,. \tag{2.2.16}$$

Now, for each fixed $m < k'$,

$$\lim_{k' \to \infty} \inf_{k \ge k'} \bar{\tau}_m^k \ge \bar{\tau}_m^\infty.$$

Finally we take $m \to \infty$ and since $\bar{\tau}_m^\infty \to \infty$ we can conclude that (2.2.16) holds for all $t \in I$. The last statement of the theorem follows from Corollary 2.2.12. □

2.2.3 Examples

Example 2.2.14 (Integrated Lyapunov condition). Consider the McKean–Vlasov stochastic differential equation (2.1.4) i.e.

$$dx_t = -x_t\left[\int_{\mathbb{R}} y^4 \mathscr{L}(x_t)(dy)\right] dt + \frac{1}{\sqrt{2}} x_t\, dw_t\,, \quad x_0 = \xi > 0\,.$$

Then for $v(x) = x^4$ we have,

$$L(x,\mu)v(x) = 3x^4 - 4x^4 \int_{\mathbb{R}} y^4\, \mu(dy)\,.$$

We see that the stronger Lyapunov condition (2.2.2) will not hold with $m_1 < 0$ (at least for chosen v, which seems to be a natural choice) and $D_k = (-k, k)$. However, integrating leads to

$$\int_{\mathbb{R}} L(x,\mu)v(x)\mu(dx) = 3\int_{\mathbb{R}} x^4 \mu(dx) - 4\left(\int_{\mathbb{R}} x^4 \mu(dx)\right)^2$$

using this we will show that the integrated Lyapunov condition (2.2.7) holds i.e. that

$$\int_{\mathbb{R}} L(x,\mu)v(x)\mu(dx) \le -\int_{\mathbb{R}} v(x)\mu(dx) + 1$$

is satisfied. To see this we note that $3a - 4a^2 \le 1 - a$ since $-1 + 4a - 4a^2 \le -(1-2a)^2$. Moreover, Assumption 2.2.7 is satisfied. Condition (2.1.6) allows us to obtain uniform-in-time integrability properties for (x_t) needed to study ergodic properties.

Example 2.2.15 (Non-linear dependence of measure and integrated Lyapunov condition). Consider the McKean–Vlasov stochastic differential equation (2.1.7) i.e.

$$dx_t = -\left(\int_{\mathbb{R}} (x_t - \alpha y)\mathscr{L}(x_t)(dy)\right)^3 dt + \left(\int_{\mathbb{R}} (x_t - \alpha y)\mathscr{L}(x_t)(dy)\right)^2 \sigma\, dw_t\,,$$

for $t \in I$ and with $x_0 \in L^4(\mathcal{F}_0, \mathbb{R})$. Assume that $m := -(6\sigma^2 - 4 + 4\alpha) > 0$. The diffusion generator given by (2.2.1) is

$$(L^\mu v)(x,\mu) = \Bigg(\frac{\sigma^2}{2}\Big(\int_{\mathbb{R}}(x-\alpha y)\mu(dy)\Big)^4 \partial_x^2 v - \Big(\int_{\mathbb{R}}(x-\alpha y)\mu(dy)\Big)^3 \partial_x v\Bigg)(x,\mu)$$
$$+ \int_{\mathbb{R}} \Bigg(\frac{\sigma^2}{2}\Big(\int_{\mathbb{R}}(z-\alpha y)\mu(dy)\Big)^4 (\partial_z\partial_\mu v)(t,x,\mu)(z)$$
$$- \Big(\int_{\mathbb{R}}(z-\alpha y)\mu(dy)\Big)^3 (\partial_\mu v)(t,x,\mu)(z)\Bigg)\mu(dz)\,.$$

We will show that for the Lyapunov function

$$v(x,\mu) = \Big(\int_{\mathbb{R}}(x-\alpha y)\mu(dy)\Big)^4,$$

we have

$$\int_{\mathbb{R}}(L^\mu v)(x,\mu)\,\mu(dx) \le m - m\int_{\mathbb{R}} v(x,\mu)\,\mu(dx)\,.$$

Indeed,

$$\partial_x v(x,\mu) = 4\Big(\int_{\mathbb{R}}(x-\alpha y)\mu(dy)\Big)^3, \quad \partial_x^2 v(x,\mu) = 12\Big(\int_{\mathbb{R}}(x-\alpha y)\mu(dy)\Big)^2,$$

$$\partial_\mu v(x,\mu)(z) = -4\alpha\Big(\int_{\mathbb{R}}(x-\alpha y)\mu(dy)\Big)^3 \text{ and } \partial_z\partial_\mu v(x,\mu)(z) = 0\,.$$

Hence

$$(L^\mu v)(x,\mu) = (6\sigma^2 - 4)\Big(\int_{\mathbb{R}}(x-\alpha y)\mu(dy)\Big)^6$$
$$+ 4\alpha\int_{\mathbb{R}}\Bigg[\Big(\int_{\mathbb{R}}(z-\alpha y)\mu(dy)\Big)^3\Big(\int_{\mathbb{R}}(x-\alpha y)\mu(dy)\Big)^3\Bigg]\mu(dz)\,.$$

Since we want an estimate over the integral of the diffusion generator we observe that

$$I := \int_{\mathbb{R}}\int_{\mathbb{R}}\Bigg[\Big(\int_{\mathbb{R}}(z-\alpha y)\,\mu(dy)\Big)^3\Big(\int_{\mathbb{R}}(x-\alpha y)\mu(dy)\Big)^3\Bigg]\mu(dz)\,\mu(dx)$$
$$\le \Bigg(\int_{\mathbb{R}}\Big|\int_{\mathbb{R}}(z-\alpha y)\,\mu(dy)\Big|^3\mu(dz)\Bigg)^2.$$

By the Cauchy–Schwarz inequality we obtain

$$I \le \int_{\mathbb{R}}\Big(\int_{\mathbb{R}}(x-\alpha y)\,\mu(dy)\Big)^6\mu(dx)\,.$$

Hence, recalling $m := -(6\sigma^2 - 4 + 4\alpha) > 0$ and using the inequality $-x^6 \leq 1 - x^4$, we obtain that

$$\int_{\mathbb{R}} (L^{\mu} v)(x,\mu)\,\mu(dx) \leq \int_{\mathbb{R}} (6\sigma^2 - 4 + 4\alpha)\left(\int_{\mathbb{R}} (x - \alpha y)\,\mu(dy)\right)^6 \mu(dx)$$
$$\leq m - m\int_{\mathbb{R}} v(x,\mu)\,\mu(dx)\,.$$

Furthermore, Assumption 2.2.7 is readily satisfied.

Example 2.2.16 (Dependence on measure not through its moments)**.** Let μ be probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and let $F_\mu^{-1} : [0,1] \to \mathbb{R}$ be the generalised inverse cumulative distribution function for this law. Recall that the α-Quantile is given by

$$F_\mu^{-1}(\alpha) := \inf\{x \in \mathbb{R} \,:\, \mu((-\infty, x]) \geq \alpha]\}\,.$$

Define the Expected Shortfall of μ at level α, as $ES_\mu(\alpha)$, as

$$ES_\mu(\alpha) := \frac{1}{\alpha}\int_0^\alpha F_\mu^{-1}(s)\,ds\,.$$

Note that for fixed α, Expected Shortfall is a Lipschitz continuous function of measure w.r.t p-th Wasserstein distances for $p \geq 1$. Indeed fix μ, $\nu \in \mathcal{P}_p(\mathbb{R})$ and observe that

$$|ES_\mu(\alpha) - ES_\nu(\alpha)| \leq \frac{1}{\alpha}\int_0^\alpha |F_\mu^{-1}(s) - F_\nu^{-1}(s)|\,ds$$
$$\leq \frac{1}{\alpha}\int_0^1 |F_\mu^{-1}(s) - F_\nu^{-1}(s)|\,ds = \frac{1}{\alpha}W_1(\mu,\nu) \leq \frac{1}{\alpha}W_p(\mu,\nu)\,,$$

where the equality above follows from [99].

We consider the following one-dimensional example, based loosely on transformed CIR:

$$dx_t = \frac{\kappa}{2}\big[((ES_{\mathscr{L}(x_t)}(\alpha) \vee \theta) - \frac{\sigma^2}{4\kappa})x_t^{-1} - x_t\big]\,dt + \frac{1}{2}\sigma\,dw_t.$$

Here x_0 satisfies $\mathbb{P}[x_0 > 0] = 1$ and $\kappa\theta \geq \sigma^2$.

Note that by defining $D := (0,\infty)$ and $D_k := (\frac{1}{k}, k)$, we have boundedness of the coefficients on D_k and from the above observations and assumptions one can easily verify that the conditions of Theorem 2.2.10 are satisfied. In particular consider $v(x) = x^2 + x^{-2}$. Then,

$$L(x,\mu)v(x) = \frac{\kappa}{2}\big[((ES_\mu(\alpha) \vee \theta) - \frac{\sigma^2}{4\kappa})x^{-1} - x\big](2(x - x^{-3})) + \frac{1}{8}\sigma^2(2 + 6x^{-4})$$
$$= \kappa\big[((ES_\mu(\alpha) \vee \theta) - \frac{\sigma^2}{4\kappa})\big] - \kappa x^2 - \big[\kappa(ES_\mu(\alpha) \vee \theta) - \sigma^2\big]x^{-4}$$
$$+ \kappa x^{-2} + \frac{\sigma^2}{4}$$
$$\leq |\kappa||ES_\mu(\alpha)| + \kappa\theta + \kappa x^{-2}$$
$$\leq |\kappa||\int_{\mathbb{R}} x\,\mu(dx)| + \kappa\theta + \kappa x^{-2}$$
$$\leq \frac{1}{2}\kappa^2 + \frac{1}{2}\int_{\mathbb{R}} x^2\,\mu(dx) + \kappa\theta + \kappa x^{-2}.$$

Integrating with respect to μ we see that condition (2.2.7) holds. Therefore, due to Theorem 2.2.10, a weak solution to the above McKean–Vlasov equation exists.

2.3 Uniqueness

In this Section we prove continuous dependence on initial conditions and uniqueness under two types of Lyapunov conditions. For the novel integrated global Lyapunov condition we provide an example that has been inspired by the work of [90] on non-uniqueness of solutions to McKean–Vlasov SDEs.

2.3.1 Assumptions and Results

Recall that $\Pi(\mu,\nu)$ denotes the set of couplings between measures μ and ν. In this section we work with a subclass of Lyapunov functions $\bar{v}\in C^2(\mathbb{R}^d)$ that has the properties: $\bar{v}\geq 0$, $\mathrm{Ker}\;\bar{v}=\{0\}$ and for all $z\in\mathbb{R}^d$ we have $\bar{v}(z)=\bar{v}(-z)$. For this class of Lyapunov functions we define a 'Wasserstein semi-distance' on $\mathcal{P}(D)$ as,

$$W_{\bar{v}}(\mu,\nu):=\left(\inf_{\pi\in\Pi(\mu,\nu)}\int_{D\times D}\bar{v}(x-y)\,\pi(dx,dy)\right). \tag{2.3.1}$$

Indeed W_v is a semi-metric and the triangle inequality does not hold in general. Note that $\bar{v}$ does not depend on a measure. For $(t,x,y)\in I\times D\times D$, $(\mu,\nu)\in\mathcal{P}(D)\times\mathcal{P}(D)$, and any $\varphi\in C^2(\mathbb{R}^d)$, we define

$$\begin{aligned}L(t,x,y,\mu,&\nu)\varphi(x-y)\\ :=&(b(t,x,\mu)-b(t,y,\nu))\partial_x\varphi(x-y)\\ &+\frac{1}{2}\mathrm{tr}\big((\sigma(t,x,\mu)-\sigma(t,y,\nu))(\sigma(t,x,\mu)-\sigma(t,y,\nu))^T\partial_x^2\varphi(x-y)\big).\end{aligned}$$

Assumption 2.3.1 (Global Lyapunov Condition). There exists $\bar{v}\in C^2(\mathbb{R}^d)$ and locally integrable, non-random, functions $g=g(t)$ and $h=h(t)$ on I, such that for any two solutions $(x_t)_{t\in I}$ and $(y_t)_{t\in I}$ to (2.1.1), with $\mathscr{L}(x_t)=\mu_t$ and $\mathscr{L}(y_t)=\nu_t$, for all $t\in I$,

$$L(t,x_t,y_t,\mu_t,\nu_t)\bar{v}(x_t-y_t)\leq g(t)\bar{v}(x_t-y_t)+h(t)W_{\bar{v}}(\mu_t,\nu_t). \tag{2.3.2}$$

Assumption 2.3.2 (Integrated Global Lyapunov Condition). There exists $\bar{v}\in C^2(\mathbb{R}^d)$ and a locally integrable, non-random function $h=h(t)$ on I, such that for any two solutions $(x_t)_{t\in I}$ and $(y_t)_{t\in I}$ to (2.1.1), with $\mathscr{L}(x_t)=\mu_t$ and $\mathscr{L}(y_t)=\nu_t$, for all $t\in I$, and for all couplings $\pi\in\Pi(\mu_t,\nu_t)$

$$\int_{D\times D}L(t,p,q,\mu_t,\nu_t)\bar{v}(p-q)\pi(dp,dq)\leq h(t)\int_{D\times D}\bar{v}(p-q)\,\pi(dp,dq). \tag{2.3.3}$$

It can be shown that the three examples given in the previous section satisfy Assumption 2.3.2 with $\bar{v}(z)=z^2$.

Theorem 2.3.3 gives a stability estimate for the solution to (2.1.1) with respect to initial condition (continuous dependence on the initial conditions).

Theorem 2.3.3 (Continuous Dependence on Initial Condition). *Let Assumption 2.2.6 hold. Let x^i, $i=1,2$ be two solutions to (2.1.1) on the same probability space such that $\mathbb{E}\bar{v}(x_0^1 - x_0^2) < \infty$.*

i) If Assumption 2.3.1 holds then for all $t \in I$

$$\mathbb{E}[\bar{v}(x_t^1 - x_t^2)] \leq \exp\left(\int_0^t [g(s) + h(s) + |h(s)|]\, ds\right) \mathbb{E}[\bar{v}(x_0^1 - x_0^2)]\,. \quad (2.3.4)$$

ii) If Assumptions 2.3.2 and either Assumption 2.2.1a, 2.2.1b or 2.2.2 hold and if there are p, q with $1/p + 1/q = 1$ and a constant κ such that for all (t, x, μ) in $I \times D \times \mathcal{P}(D)$

$$|\partial_x \bar{v}(x-y)|^{2p} + |\sigma(t,x,\mu)|^{2q} + |\sigma(t,y,\nu)|^{2q} \leq \kappa(1 + v(t,x,\mu) + v(t,y,\nu)) \quad (2.3.5)$$

then for all $t \in I$

$$\mathbb{E}[\bar{v}(x_t^1 - x_t^2)] \leq \exp\left(\int_0^t h(s)\, ds\right) \mathbb{E}[\bar{v}(x_0^1 - x_0^2)]\,. \quad (2.3.6)$$

First we note that in the case when I is a finite time interval then the sign of the functions g and h plays no significant rôle. In relation to the study of ergodic SDEs e.g. (18) in [15] we make the following observations. If $I = [0, \infty)$ and Assumption 2.3.1 holds with $g + h + |h| < 0$ then $\lim_{t\to\infty} \mathbb{E}\bar{v}(x_t^1 - x_t^2) = 0$. However we see that while the spatial dependence of coefficients can play a positive rôle for the stability of the equation (if g is negative) it seems that the measure dependence never has such a positive rôle, regardless of the sign of h. If $I = [0, \infty)$ and we are in the second case of Theorem 2.3.3 then negative h may contribute to stability (but unlike the first case we also need the condition (2.3.5)).

Proof. Note that if we are in case ii) then, in the following we set $g = 0$ for all $t \in I$. Let

$$\varphi(t) = \exp\left(-\int_0^t [g(s) + h(s)]\, ds\right)\,.$$

Applying the classical Itô formula to $\varphi\, \bar{v}(x^1 - x^2)$ we have that for $t \in I$

$$\begin{aligned}
&\varphi(t)\bar{v}(x_t^1 - x_t^2) \\
&= \bar{v}(x_0^1 - x_0^2) \\
&+ \int_0^t \varphi(s)\big[L(s, x_s^1, x_s^2, \mathscr{L}(x_s^1), \mathscr{L}(x_s^2))\bar{v}(t, x_s^2 - x_s^2) - (g(s) + h(s))\bar{v}(x_s^1 - x_s^2)\big]\, ds \\
&+ \int_0^t \varphi(s)\partial_x \bar{v}(x_s^1 - x_s^2)(\sigma(s, x_s^1, \mathscr{L}(x_s^1)) - \sigma(s, x_s^2, \mathscr{L}(x_s^2)))dw_s.
\end{aligned} \quad (2.3.7)$$

Case i) Assumption 2.3.1 implies

$$\begin{aligned}
\varphi(t)\bar{v}(x_t^1 - x_t^2) \leq \bar{v}(x_0^1 - x_0^2)
&+ \int_0^t \varphi(s)\big[h(s)W_{\bar{v}}(\mathscr{L}(x_s^1), \mathscr{L}(x_s^2)) - h(s)\bar{v}(x_s^1 - x_s^2)\big]\, ds \\
&+ \int_0^t \varphi(s)\partial_x \bar{v}(x_s^1 - x_s^2)(\sigma(s, x_s^1, \mathscr{L}(x_s^1)) - \sigma(s, x_s^2, \mathscr{L}(x_s^2)))dw_s.
\end{aligned}$$

Define the stopping times $\{\tau_m^i\}_{m\geq 1}$, $i=1,2$ and $\{\tau_m\}_{m\geq 1}$

$$\tau_m^i := \inf\{t\in I\,:\, x_t^i \notin D_m\}\,,\ \ i=1,2 \ \text{ and }\ \tau_m := \tau_m^1\wedge\tau_m^2\,.$$

By Definition 2.2.8 we know that $x^i\in C(I;D)$ a.s. and so $\tau_m^i\nearrow\infty$ a.s. and hence $\tau_m\nearrow\infty$ a.s. as $m\to\infty$. The local boundedness of σ ensures that the stochastic integral in the above is a martingale on $[t\wedge\tau_m]$, hence

$$\begin{aligned}
&\mathbb{E}[\varphi(t\wedge\tau_m)\bar v(x^1_{t\wedge\tau_m}-x^2_{t\wedge\tau_m})]\\
&\leq \mathbb{E}[\bar v(x_0^1-x_0^2)]+\mathbb{E}\left[\int_0^{t\wedge\tau_m}\varphi(s)\big[h(s)W_{\bar v}(\mathscr{L}(x_s^1),\mathscr{L}(x_s^2))-h(s)\bar v(x_s^1-x_s^2)\big]\,ds\right]\\
&\leq \mathbb{E}[\bar v(x_0^1-x_0^2)]+\mathbb{E}\left[\int_0^{t}\varphi(s)\big[|h(s)|\bar v(x_s^1-x_s^2)\big]\,ds\right]\,,
\end{aligned}$$

where the last inequality follows from the definition of the semi-Wasserstein distance. Since $\tau_m\nearrow\infty$ as $m\to\infty$, application of Fatou's Lemma gives

$$\mathbb{E}[\varphi(t)\bar v(x_t^1-x_t^2)]\leq \mathbb{E}[\bar v(x_0^1-x_0^2)]+\int_0^t |h(s)|\mathbb{E}[\varphi(s)\bar v(x_s^1-x_s^2)]\,ds.$$

From Gronwall's lemma we get (2.3.4).

Case ii) Taking expectation in (2.3.7), recalling that in this case $g=0$ and then using Assumption 2.3.2 we have

$$\begin{aligned}
&\mathbb{E}\left[\varphi(t)\bar v(x_t^1-x_t^2)\right]\\
&\leq \mathbb{E}[\bar v(x_0^1-x_0^2)]+\mathbb{E}\left[\int_0^t\varphi(s)\partial_x\bar v(x_s^1-x_s^2)(\sigma(s,x_s^1,\mathscr{L}(x_s^1))-\sigma(s,x_s^2,\mathscr{L}(x_s^2)))\,dw_s\right].
\end{aligned}$$

Corollary 2.2.12 together with (2.3.5) and local integrability of g and h ensures that stochastic integral in the above expression is a martingale. Indeed

$$\begin{aligned}
&\int_0^t\varphi(s)^2\mathbb{E}\left[|\partial_x\bar v(x_s^1-x_s^2)|^2|\sigma(s,x_s^1,\mathscr{L}(x_s^1))-\sigma(s,x_s^2,\mathscr{L}(x_s^2))|^2\right]\,ds\\
&\leq \int_0^t\varphi(s)^2\mathbb{E}\left[\frac{1}{p}|\partial_x\bar v(x_s^1-x_s^2)|^{2p}+\frac{1}{q}|\sigma(s,x_s^1,\mathscr{L}(x_s^1))-\sigma(s,x_s^2,\mathscr{L}(x_s^2))|^{2q}\right]\,ds\\
&\leq c_{p,q}\int_0^t\varphi(s)^2\mathbb{E}\left[\partial_x\bar v(x_s^1-x_s^2)|^{2p}+|\sigma(s,x_s^1,\mathscr{L}(x_s^1))|^{2q}+|\sigma(s,x_s^2,\mathscr{L}(x_s^2))|^{2q}\right]\,ds\\
&\leq c_{p,q}\int_0^t\varphi(s)^2\kappa\left(1+\mathbb{E}[v(s,x_s^1,\mathscr{L}(x_s^1))]+\mathbb{E}[v(s,x_s^2,\mathscr{L}(x_s^2))]\right)\,ds<\infty\,.
\end{aligned}$$

Hence

$$\varphi(t)\mathbb{E}[\bar v(x_t^1-x_t^2)]\leq \mathbb{E}[\bar v(x_0^1-x_0^2)]\,.$$

□

Corollary 2.3.4. *Let the conditions for either case i) or ii) of Theorem 2.3.3 hold with either* $\sup_{t\in I}\exp\left(\int_0^t[g(s)+h(s)+|h(s)|]\,ds\right)<\infty$ *or* $\sup_{t\in I}\exp\left(\int_0^t h(s)\,ds\right)<\infty$ *respectively. If $x_0^1=x_0^2$ a.s. then the solutions to (2.1.1) are pathwise unique.*

Proof. Since Ker $\bar v=\{0\}$, we have that for all $t\in I$, $\mathbb{P}(x_t^1=x_t^2)=1$. Then, since the processes have continuous paths, we can conclude that they are indistinguishable.

□

2.3.2 Example due to Scheutzow

Consider the McKean–Vlasov SDE of the form

$$x_t = x_0 + \int_0^t B(x_s, \mathbb{E}[\bar{b}(x_s)])\, ds + \int_0^t \Sigma(x_s, \mathbb{E}[\bar{\sigma}(x_s)])\, dw_s\,. \tag{2.3.8}$$

Our study of this more specific form of McKean–Vlasov SDE is inspired by [90], where it has been shown that in the case when $\Sigma = 0$ and either of functions B or $\bar{b}$ is only locally Lipschitz continuous then uniqueness, in general, does not hold. We will show that if we impose some structure on the local behaviour of the functions then these, together with the integrability conditions established in Theorem 2.2.10, are enough to obtain unique solution (2.3.8). To be more specific: we impose a local (in the second variable) monotone condition on functions B and Σ, which is weaker than the local (in the second variable) Lipschitz condition, and local Lipschitz conditions on functions $\bar{b}$ and $\bar{\sigma}$.

Assumption 2.3.5.

i) Local Monotone condition:
there exists a locally bounded function $M = M(x', y', x'', y'')$ such that $\forall x, x'$, x'', y, y', $y'' \in D$

$$\begin{aligned}&2(x-y)(B(x,x') - B(y,y')) + |\Sigma(x,x'') - \Sigma(y,y'')|^2 \\ &\qquad\qquad \leq M(x',y',x'',y'')(|x-y|^2 + |x'-y'|^2 + |x''-y''|^2)\end{aligned}$$

There exists a constant κ such that:

ii) $\forall (t,x,\mu) \in I \times D \times \mathcal{P}(D)$ $|\bar{b}(x)| + |\bar{\sigma}(x)| \leq \kappa(1 + v(t,x,\mu))$, and

iii) $\forall (t,x,y,\mu) \in I \times D \times D \times \mathcal{P}(D)$

$$|\bar{b}(x) - \bar{b}(y)| + |\bar{\sigma}(x) - \bar{\sigma}(y)| \leq \kappa(1 + \sqrt{v(t,x,\mu)} + \sqrt{v(t,y,\mu)})|x-y|\,.$$

Theorem 2.3.6. *If Assumptions 2.2.2 hold, if $\sup_{t\in I} M(t) < \infty$ and if Assumptions 2.2.6, 2.3.5 hold then the solution to (2.3.8) is unique.*

We will need the following observation: if $\pi \in \Pi(\mu,\nu)$ then, due to the theorem on disintegration, (see for example [2, Theorem 5.3.1]) there exists a family $(P_x)_{x\in D} \subset \mathcal{P}(D)$ such that

$$\int_{D\times D} f(x,y)\,\pi(dx,dy) = \int_D \left(\int_D f(x,y)\,P_x(dy)\right)\,\mu(dx)$$

for any $f = f(x,y)$ which is a π-integrable function on $D \times D$. In particular if $f = f(x)$ then

$$\int_{D\times D} f(x)\,\pi(dx,dy) = \int_D f(x)\left(\int_D P_x(dy)\right)\,\mu(dx) = \int_D f(x)\,\mu(dx)\,.$$

Proof. Our aim is to show that Assumption 2.3.1 holds, since uniqueness would follow from Corollary 2.3.4. We know, from Lemma 2.2.11 that for any $t \in I$ we have the estimate
$\int_D v(t,x,\mathscr{L}(x_t))\mathscr{L}(x_t)(dx) \leq \sup_{t\in I} M(t)$ and so it suffices to verify (2.3.2) for measures μ such that $\int_D v(t,x,\mu)\,\mu(dx) \leq \sup_{t\in I} M(t)$. From Assumption 2.3.5 i), we have

$$\begin{aligned}
&2(x-y)(B(x,\mu)-B(y,\nu)) + |\Sigma(x,\mu)-\Sigma(y,\nu)|^2 \\
&\qquad\qquad \leq M(x',y',x'',y'')[|x-y|^2 + |x'-y'|^2 + |x''-y''|^2],
\end{aligned}$$

where $x' = \int_D \bar{b}(z)\mu(dz)$, $y' = \int_D \bar{b}(z)\nu(dz)$, $x'' = \int_D \bar{\sigma}(z)\mu(dz)$ and $y'' = \int_D \bar{\sigma}(z)\nu(dz)$. We note that each of $|x'|$,$|y'|$,$|x''|$ and $|y''|$ are in a compact subset of $\mathbb{R}$, due to Assumption 2.3.5 ii), since

$$\kappa\left(1+\int_D v(t,z,\mu)\,\mu(dz)\right) + \kappa\left(1+\int_D v(t,z,\mu)\,\nu(dz)\right) \leq 2\kappa(1+\sup_{t\in I} M(t)).$$

As M maps bounded sets to bounded sets we can choose a constant g sufficiently large so that $M(x',y',x'',y'') \leq g$ for all μ,ν.

We apply the remark on disintegration to see that

$$|x'-y'|^2 = \left|\int_D \bar{b}(\bar{x})\mu(d\bar{x}) - \int_D \bar{b}(\bar{y})\nu(d\bar{y})\right|^2 = \left|\int_{D\times D} (\bar{b}(\bar{x})-\bar{b}(\bar{y}))\,\pi(dx,d\bar{y})\right|^2.$$

From Assumption 2.3.5 iii), we get

$$\begin{aligned}
&|x'-y'|^2 \\
&\leq \kappa^2 \int_{D\times D} (1+\sqrt{v(t,\bar{x},\mu)}+\sqrt{v(t,\bar{y},\mu)})^2\,\pi(d\bar{x},d\bar{y}) \int_{D\times D} |\bar{x}-\bar{y}|^2\pi(d\bar{x},d\bar{y}) \\
&\leq 3\kappa^2(1+2\sup_{t\in I} M(t)) \int_{D\times D} |\bar{x}-\bar{y}|^2\pi(d\bar{x},d\bar{y}).
\end{aligned}$$

Since the calculation for $|x''-y''|^2$ is identical we finally obtain

$$\begin{aligned}
&2(x-y)(B(x,\mu)-B(y,\nu)) + |\Sigma(x,\mu)-\Sigma(y,\nu)|^2 \\
&\qquad\qquad \leq g|x-y|^2 + 6g\kappa^2(1+2\sup_{t\in I} M(t)) \int_{D\times D} |\bar{x}-\bar{y}|^2\pi(d\bar{x},d\bar{y})
\end{aligned}$$

as required to have Assumption 2.3.1 satisfied with $\bar{v}(z) = |z|^2$. □

2.4 Invariant Measures

We will establish the existence of a stationary measure for semigroups on $C_b(\mathcal{P}_2(D))$ associated with the flow of laws of solutions to (2.1.1) where the coefficients b and σ do not depend on t, via the Krylov–Bogolyubov Theorem (see [86, Chapter 7]). One cannot consider a semigroup acting on $C_b(D)$ due to the measure-dependence of the coefficients. Let the conditions of Theorem 2.2.10 hold with suitable assumptions on m_1 and m_2 such that we are within the regime where $I = [0,\infty)$.

Define the semigroup $(\mathscr{P}_t)_{t\geq 0}$ by

$$\mathscr{P}_t\phi(\mu) = \phi(\mathscr{L}(x_t^\mu)) \text{ for } \phi \in C_b(\mathcal{P}_2(D)) \text{ and } t \geq 0. \tag{2.4.1}$$

Here x_t^μ denotes a solution to (2.1.1) started from $x_0 \sim \mu$. To ensure that $\mathscr{L}(x_t^\mu) \in \mathcal{P}_2(D)$ we assume that the conditions of Theorem 2.2.10 hold with V satisfying $V(t,x) \geq |x|^2$. If $D = \mathbb{R}^d$ then we can apply the chain rule for functions of measures from e.g. [19] or [27] to obtain that for $\phi \in \mathcal{C}^{(1,1)}(\mathcal{P}_2(D))$

$$\begin{aligned}&\phi(\mathscr{L}(x_t)) - \phi(\mathscr{L}(x_0))\\&= \int_0^t \langle \mathscr{L}(x_s^\mu), b(\cdot, \mathscr{L}(x_s^\mu))\partial_\mu\phi(\mathscr{L}(x_s^\mu)) + \operatorname{tr}[a(\cdot, \mathscr{L}(x_s^\mu))\partial_y\partial_\mu\phi(\mathscr{L}(x_s^\mu))]\rangle\, ds.\end{aligned} \tag{2.4.2}$$

In the case that $D \subset \mathbb{R}^d$ we have to assume that there is $k \in \mathbb{N}$ such that $V(t,x) \geq |x|^2$ for $x \in D \setminus D_k$. We consider first $x^{k,\mu}$ given by (2.2.11) started from μ^k (μ restricted to D_k with external mass moved to 0). By Proposition A.3.2 we have for $\phi \in \mathcal{C}^{(1,1)}(\mathcal{P}_2(D))$ that

$$\begin{aligned}\phi(\mathscr{L}(x_t^k)) - \phi(\mathscr{L}(x_0^k)) = \int_0^t \Big\langle &\mathscr{L}(x_s^{k,\mu}), b(\cdot, \mathscr{L}(x_s^{k,\mu}))\partial_\mu\phi(\mathscr{L}(x_s^{k,\mu}))\\&+ \operatorname{tr}\Big[a(\cdot, \mathscr{L}(x_s^{k,\mu}))\partial_y\partial_\mu\phi(\mathscr{L}(x_s^{k,\mu}))\Big]\Big\rangle\, ds.\end{aligned} \tag{2.4.3}$$

From Lemma 2.2.11 we get that $\sup_k \sup_t \mathbb{E}|x_t^k|^2 < \infty$. Moreover Lemma 2.2.13 implies, together with Prohorov's theorem, convergence of a subsequence of the laws (and since we know the limit of these is given by (2.1.1) due to the proof of Theorem 2.2.10). We thus have $W_2(\mathscr{L}(x_t^k), \mathscr{L}(x_t)) \to 0$ as $k \to \infty$. Due to continuity of coefficients b, σ and since $\phi \in \mathcal{C}^{(1,1)}(\mathcal{P}_2(D))$ we can take the limit $k \to \infty$ in (2.4.3) to obtain (2.4.2).

The conditions for Krylov-Bogolyubov's theorem to hold are that the Markov semigroup is Feller and a tightness condition. As we are not assuming any non-degeneracy of the diffusion coefficient we cannot always guarantee that the semigroup is Feller. See, however, Lemma 2.4.2 for a partial result.

Theorem 2.4.1. *Let the conditions of Theorem 2.2.10 hold with $I = [0,\infty)$, and $V(t,x) \geq |x|^2$ for $x \in D \setminus D_k$ for some $k \in \mathbb{N}$. If the semigroup $(\mathscr{P}_t)_{t\geq 0}$ given by (2.4.1) is Feller then there exists an invariant measure.*

Proof of Theorem 2.4.1. Fix $\mu \in \mathcal{P}_2(D)$ and let x^μ be a solution to (2.2.16). Setting $\pi_t(\mu, B) := \delta_{\mathscr{L}(x_t^\mu)}(B)$, then from (2.4.1), we have that

$$\mathscr{P}_t\phi(\mu) = \phi(\mathscr{L}(x_t^\mu)) = \int_{\mathcal{P}(D)} \delta_{\mathscr{L}(x_t^\mu)}(\nu)\phi(\nu)\, d\nu = \int_{\mathcal{P}(D)} \phi(\nu)\, \pi_t(\mu, d\nu)\,.$$

Define the family of measures $(m_t^\mu)_{t\geq 0} \subset \mathcal{P}(\mathcal{P}(D))$ by

$$m_t^\mu(B) := \frac{1}{t}\int_0^t \pi_s(\mu, B)\, ds = \frac{1}{t}\int_0^t \delta_{\mathscr{L}(x_s^\mu)}(B)\, ds\,, \quad B \in \mathcal{B}(\mathcal{P}_2(D))\,.$$

To apply the Krylov–Bogolyubov Theorem we need to show that the family $(m_t^\mu)_{t\geq 0}$ is tight. We observe that for all $f \in B(D)$ we have,

$$\langle I(m_t^\mu), f\rangle = \int_{\mathcal{P}(D)} \langle \nu, f\rangle \, m_t^\mu(d\nu) = = \frac{1}{t}\int_0^t \langle \mathscr{L}(x_s^\mu), f\rangle ds$$

By Proposition 1.5.2ii), it remains to show that family of intensity measures $(I(m_t))_{t\geq 0} \subset \mathcal{P}(D)$ is tight. For $B \in \mathcal{B}(D)$ we have

$$I(m_t^\mu)(B) = \frac{1}{t}\int_0^t \langle \mathscr{L}(x_s^\mu), \mathbb{1}_B\rangle \, ds = \frac{1}{t}\int_0^t \mathscr{L}(x_s^\mu)(B)\, ds = \frac{1}{t}\int_0^t \mathbb{P}(x_s^\mu \in B)\, ds\,.$$

By Fatou's Lemma and Lemma 2.2.13 we know that for any $\varepsilon > 0$ there exists sufficiently large m_0 such that for all $m > m_0$ we have $\sup_{t\in I}\mathbb{P}[x_t^\mu \notin D_m] < \varepsilon$. Therefore $I(m_t^\mu)(D\setminus D_m) = \frac{1}{t}\int_0^t \mathbb{P}(x_s^\mu \notin D_m)ds < \varepsilon$ and hence $(I(m_t^\mu))_{t\geq 0}$ is tight. □

Lemma 2.4.2. *Let the assumptions of Theorem 2.2.10 hold for $I = [0,\infty)$ along with either Assumption 2.3.1 or 2.3.2. Assume further that*

$$W_{\bar v}(\mu,\nu) < \infty \quad \textit{for } \mu,\nu \textit{ in} \quad \mathcal{P}_v(D) := \left\{\mu \in \mathcal{P}(D) : \int_D v(0,x,\mu)\,\mu(dx) < \infty\right\}.$$

Then the semigroup $(\mathscr{P}_t)_{t\geq 0}$ acting on $C_b(\mathcal{P}_v(D))$ and defined as in (2.4.1) is Feller.

Note that here, we are considering a semigroup acting on space of measures possibly different to that previously considered. In the case where v and $\bar v$ are polynomials, one may replace the assumption of the Feller property in Theorem 2.4.1 with the assumptions of Lemma 2.4.2 and $W_{\bar v}(\mu,\nu) < \infty$ for any $\mu,\nu \in \mathcal{P}_v(D)$ is no longer required.

Proof. Fix $t \in I$ and $\mu_1,\mu_2 \in \mathcal{P}_v(D)$. From the continuous dependence on initial condition, Theorem 2.3.3, we have

$$\begin{aligned} W_{\bar v}(\mathscr{L}(x_t^{\mu_1}), \mathscr{L}(x_t^{\mu_2})) \leq \mathbb{E}[\bar v(x_t^{\mu_1} - x_t^{\mu_2})] &\leq c_t \mathbb{E}[\bar v(x_0^{\mu_1} - x_0^{\mu_2})] \\ &= c_t \int_{D\times D} \bar v(x-y)\pi(dx,dy)\,. \end{aligned}$$

Taking the infimum over all the possible couplings yields,

$$W_{\bar v}(\mathscr{L}(x_t^{\mu_1}), \mathscr{L}(x_t^{\mu_2})) \leq c_t W_{\bar v}(\mu_1,\mu_2). \tag{2.4.4}$$

Let $\varepsilon > 0$ be given. For any $\phi \in C_b(\mathcal{P}_v(D))$ and $\mu \in \mathcal{P}_v(D)$ there is $\delta_{\phi,\mu}$ such that $W_{\bar v}(\mu,\nu) < \delta_{\phi,\mu}$ implies that $|\phi(\mu) - \phi(\nu)| < \varepsilon$. Now, by the uniqueness of solutions x^μ, for fixed time $t \in I$, $\delta_{\mu_1}(t) := c_t^{-1}\delta_{\phi,\mathscr{L}(x_t^{\mu_1})}$. Then, due to (2.4.4), if $W_{\bar v}(\mu_1,\mu_2) \leq \delta_{\mu_1}(t)$ then $W_{\bar v}(\mathscr{L}(x_t^{\mu_1}), \mathscr{L}(x_t^{\mu_2})) < \delta_{\phi,\mathscr{L}(x_t^{\mu_1})}$ and we get $|P_t\phi(\mu_1) - P_t\phi(\mu_2)| < \varepsilon$ as required. □

Chapter 3

The Conditional Propagation of Chaos and the McKean-Vlasov Dynamics with Common Noise

The previous chapter concerned the limiting behaviour of mean field interacting particle systems, where the particles were driven by independent (private) noises. In many applications however, particularly in finance, the particles'/agents' dynamics are influenced by the same random input. This may, for example, be a stock to which all agents have access. A review of related literature is found in the next chapter. The common noise to which particles will be subjected in this thesis will be modelled as a Brownian motion, labelled B.

This chapter discusses the common noise analogue of the propagation of chaos. This notion of conditional propagation of chaos is shown to hold for a particular family of particle systems, exposing some of the intricacies of the systemic noise setting.

Recall the class of interacting particle systems introduced in Section 1.6. Each system has a fixed number of particles denoted N, and the dynamics of the i^{th} particle is given by the stochastic differential equation (recall that the coefficients are progressive):

$$\begin{aligned}
X_t^{i,N} &= X_0^{i,N} + \int_0^t b(s, X^{i,N}, \mu^N)\, ds + \int_0^t \sigma(s, X^{i,N}, \mu^N)\, dW_s^{i,N} \\
&\quad + \int_0^t \rho(s, X^{i,N}, \mu^N)\, dB_s, \\
\mu_t^N &:= \frac{1}{N} \sum_{j=1}^{N} \delta_{X_{\cdot\wedge t}^{j,N}}.
\end{aligned} \tag{MFSCN}$$

The i^{th} particle within the system is driven by its private (idiosyncratic) Brownian motion $W^{i,N}$ of dimension d_W and a common (systemic) Brownian motion B of dimension d_B. To be precise, the processes $\{W^{j,N}\}_{j=1}^N$ and B are $N+1$ independent Brownian motions of dimension d_W and d_B, respectively. Additionally, the initial values $\{X^{i,N}\}_{i=1}^N$ are assumed to be exchangeable. The particles interact through the empirical measure of their paths $\mu_t^N = \frac{1}{N}\sum_{j=1}^N \delta_{X_{\cdot\wedge t}^{j,N}}$.

Intuitively one expects that, in general, the limit of the empirical measures should remain stochastic due to the influence of the common noise B. The limit is identified in the strong (Lipschitz) setting by Kurtz and Xiong [66, 67] as having corresponding McKean-Vlasov dynamics given by the following system of equations:

$$\begin{aligned} X_t =& X_0 + \int_0^t b(s,X,\mu)ds + \int_0^t \sigma(s,X,\mu)dW_s + \int_0^t \rho(s,X,\mu)dB_s \\ \mu_t =& \mathscr{L}(X_{\cdot\wedge t}|\mathcal{F}_t^B). \end{aligned} \tag{KX}$$

In this strong formulation, a solution is defined to be a complete probability space supporting independent Brownian motions B and W along with X and μ that are adapted to $\mathcal{F}^{B,W,X_0}$ satisfying (KX). In [66, 67], Kurtz and Xiong show that the unique solution to (KX) arises as the limit of finite particle systems and they provide a rate of convergence.

However, should one wish to obtain solutions via weak-convergence methods, this equation is not generally appropriate due to the implicit adaptedness condition in the equation $\mu_t = \mathscr{L}(X_{\cdot\wedge t}|\mathcal{F}_t^B)$.

Earlier, Dawson and Vaillancourt [33, 98] characterised the empirical measures' weak limit points as solutions to local martingale problems written on $C([0,\infty);\mathcal{P}(\mathbb{R}^d))$. As this characterisation is suited to weak convergence arguments, they only assume ellipticity, a weak growth condition and continuity of the coefficients. Furthermore, they demonstrate that any such solution to their local martingale problem is the law of a weak solution to a stochastic partial differential equation. This connection is also demonstrated by Kurtz and Xiong [67].

A weak form of (KX) was introduced by Lacker, Delarue and Carmona [24] in their analysis of the corresponding mean field games, by imposing a compatibility structure on their solutions. This definition is very close to the one used in this thesis. Also, Ledger and Søjmark use such a form which they call the 'relaxed problem' in [75, 76]. The notions of immersion and compatibility are recalled in the following definition. The reader is referred to [7, 26, 65, 69] for more on these concepts and Appendix B.1 for some equivalent conditions.

Definition 3.0.1 (Immersion and Compatibility)**.** Let two filtrations $\mathbb{F}$ and $\mathbb{G}$ on a probability space $(\Omega,\mathcal{F},\mathbb{P})$ be such that $\mathbb{F}\subset\mathbb{G}$. Then $\mathbb{F}$ is said to be immersed in $\mathbb{G}$ under $\mathbb{P}$ if every square integrable $\mathbb{F}$ martingale is a $\mathbb{G}$ martingale. For two stochastic processes X and Y defined on this probability space, X is said to be compatible with Y[1] if $\mathbb{F}^Y$ is immersed in $\mathbb{F}^{X,Y} := \mathbb{F}^X \vee \mathbb{F}^Y$ under $\mathbb{P}$.

3.1 McKean-Vlasov SDE with Common Noise

The following equation will be referred to as the McKean-Vlasov SDE with common noise:

$$\begin{aligned} X_t &= X_0 + \int_0^t b(s,X,\mu)\,ds + \int_0^t \sigma(s,X,\mu)\,dW_s + \int_0^t \rho(s,X,\mu)\,dB_s, \\ \mu_t &= \mathscr{L}(X_{\cdot\wedge t}|\mathcal{F}_t^{B,\mu}). \end{aligned} \tag{MKVCN}$$

[1]Strictly speaking, one should say X is compatible with Y under $\mathbb{P}$. However, no reference to the probability measure is made if the context is unambiguous.

The processes B and W are independent Brownian motions. The filtration $\{\mathcal{F}_t^{B,\mu}\}_{t\in I}$ is defined to be the natural filtration generated by the common noise B and the measure valued process μ. Let the regular conditional distribution of $X_{\cdot\wedge s}$ given $\mathcal{F}_s^{B,\mu}$ be denoted $\mathscr{L}(X_{\cdot\wedge s}|\mathcal{F}_s^{B,\mu})$. It is sensible to define a notion of a weak solution that is achievable via weak convergence methods that maintains the correspondence to the associated SPDEs. Such a definition is made as follows:

Definition 3.1.1 (Weak Solution to the McKean–Vlasov SDE with Common Noise)**.** A weak solution to the McKean-Vlasov SDE with common noise consists of a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ equipped with independent $\mathbb{F}$ Brownian motions B and W, along with $\mathbb{F}$ adapted processes X and μ that are $\mathbb{R}^{d_X}$ and $\mathcal{P}(\mathcal{C})$ valued respectively, satisfying the following conditions:

i) $\int_0^t |b(s, X_{\cdot\wedge s}, \mu_s))| + |\sigma(s, X_{\cdot\wedge s}, \mu_s))|^2 + |\rho(s, X_{\cdot\wedge s}, \mu_s))|^2\, ds < \infty$ $\mathbb{P}$-a.s. for all $t \in I$.

ii) X is compatible with (B, μ), (X, μ) is compatible with (B, W, X_0) and for $s, t \in I$ with $s \leq t$, $\sigma(W_r - W_s : s \leq r \leq t) \perp\!\!\!\perp \mathcal{F}_t^{B,\mu} \vee \mathcal{F}_s^X$.[2]

iii) The equation (MKVCN) holds $\mathbb{P}$ almost surely for all $t \in I$.

Remark 3.1.2. The first compatibility condition, in tandem with the equation $\mu_t = \mathscr{L}(X_{\cdot\wedge t}|\mathcal{F}_t^{B,\mu})$ for all $t \in I$ is equivalent to $\mu = \mathscr{L}(X_{\cdot\wedge t}|\mathcal{F}_\infty^{B,\mu})$ for all $t \in I$ - see Proposition 4.1.5.

3.2 u-Chaoticity for Non-Constant u

Before defining the conditional propagation of chaos, a definition is made that provides an analogue - appropriate to this setting - to the u-chaoticity of Sznitman (see Definition 1.5.1).

Definition 3.2.1 (u-Chaoticity (for non-constant u))**.** Let E be a Polish space, and $\{u_N\}_{N\in\mathbb{N}}$ a sequence of symmetric probability measures, each defined on a corresponding product space E^N. Then, u_N is said to be u-chaotic, for u a probability measure on $\mathcal{P}(E)$, if for $\phi_i \in C_b(E)$, and any $k \in \mathbb{N}$,

$$\lim_{N\to\infty} \langle u_N, \phi_1 \otimes \cdots \otimes \phi_k \otimes 1 \cdots \otimes 1\rangle = \int_{\mathcal{P}(E)} \prod_{i=1}^{k} \langle \mu, \phi_i\rangle u(d\mu) \tag{3.2.1}$$

With this definition, it is possible to prove a similar equivalence to that given in part i) of Proposition 1.5.2:

Proposition 3.2.2. *u_N is u-chaotic if and only if the sequence of empirical measures $\mu^N = \frac{1}{n}\sum_{i=1}^N \delta_{X^{i,N}}$ ($\mathcal{P}(E)$-valued random elements on (E^N, u_N), $X^{i,N}$ canonical coordinates on E^N) converges weakly in distribution to a random measure μ with law u.*

[2] In an attempt to aid the readability of the definition, it is intentionally tautological: the second compatibility condition is implied by the adaptedness criteria demanded earlier in the definition.

The proof of Proposition 3.2.2 uses methods adapted to distributional statements. Since the empirical measure no longer has a deterministic limit, one is unable to apply the L^2 convergence argument provided by Sznitman. If one tries to follow the proof of Proposition 1.5.2, then one needs the convergence $\mathbb{E}[\phi(X^{1,N})\langle\mu,\phi\rangle] \to \mathbb{E}[\langle\mu,\phi\rangle^2]$ for any ϕ continuous and bounded and μ a random measure with distribution u. This would require at least an assumption of stable convergence, a notion of convergence stronger than weak convergence, yet weaker than convergence in probability - see [12] section 8.10(xi) or [53]. That this argument no longer works in this setting is not too surprising since there is no longer weak convergence to a deterministic limit where one would be able to elevate to convergence in probability.

Proof of Proposition 3.2.2. **Necessity:** In order to show that $\mu^N \to \mu$ in distribution, it is enough to show that for all $\phi \in C_b(\mathcal{C})$, $\langle\mu^N,\phi\rangle \overset{w}{\to} \langle\mu,\phi\rangle$, see [56]. Fix $\phi \in C_b(\mathcal{C})$. Since ϕ is bounded, $\{\langle\mu^N,\phi\rangle\}_{N\in\mathbb{N}}$ and $\langle\mu,\phi\rangle$ are uniformly bounded and therefore may be viewed as random variables valued in a compact space $I_\phi := [-||\phi||,||\phi||]$. Then, by the Stone-Weierstrass theorem, it suffices to show that $\mathbb{E}_N[f(\langle\mu^N,\phi\rangle)] \to \mathbb{E}[f(\langle\mu,\phi\rangle)]$ for all polynomials on I_ϕ. By linearity of expectation it remains to prove the convergence for $f(x) = x^p$ for any $p \in \mathbb{N}$. Fix $p \in \mathbb{N}$ and consider $N \geq p$. Then,

$$
\begin{aligned}
\mathbb{E}_N[(\langle\mu^N,\phi\rangle)^p] =& \mathbb{E}_N\left[\left(\frac{1}{N}\sum_{i=1}^N \phi(X^{i,N})\right)^p\right] \\
=& \frac{1}{N^p}\mathbb{E}_N\left[\sum_{p_1+\dots+p_N=p,\, p_i\in\mathbb{N}_0} \binom{p}{p_1,\dots,p_N} \phi(X^{1,N})^{p_1}\cdots\phi(X^{N,N})^{p_N}\right] \\
=& \frac{1}{N^p}\left(\mathbb{E}_N\left[\sum_{p_1+\dots+p_N=p,\max_i p_i=1} p!\phi(X^{1,N})^{p_1}\cdots\phi(X^{N,N})^{p_N}\right]\right. \\
&\left. + \mathbb{E}_N\left[\sum_{p_1+\dots+p_N=p,\max_i p_i>1} \binom{p}{p_1,\dots,p_N} \phi(X^{1,N})^{p_1}\cdots\phi(X^{N,N})^{p_N}\right]\right) \\
=& \frac{1}{N^p}\left(\mathbb{E}_N\left[p!\binom{N}{p}\phi(X^{1,N})\cdots\phi(X^{p,N})\right] + o(N^p)\right) \\
=& \frac{1}{N^p}\left(N^p\mathbb{E}_N\left[\phi(X^{1,N})\cdots\phi(X^{p,N})\right] + o(N^p)\right) \\
&\to \mathbb{E}[\langle\mu,\phi\rangle^p] \quad \text{as } N\to\infty.
\end{aligned}
\tag{3.2.2}
$$

The second equality in the above holds by the multinomial theorem. Recall that

$$
\binom{p}{p_1,\dots,p_N} := \frac{p!}{p_1!\cdots p_N!}.
$$

The symbol o is the little-o Landau notation: for a positive function g, and a function f, one says that $f(N) = o(g(N))$ for $N \to \infty$ if $\lim_{N\to\infty}\frac{f(N)}{g(N)} = 0$. The fourth equality in (3.2.2) follows from the exchangeability and the following considerations.

One can estimate,

$$\begin{aligned}&\left|\mathbb{E}_N\left[\sum_{p_1+\dots+p_N=p,\max_i p_i>1}\binom{p}{p_1,\dots,p_N}\phi(X^{1,N})^{p_1}\cdots\phi(X^{N,N})^{p_N}\right]\right|\\ \leq& p!||\phi||^p\left|\left\{p_1,\dots,p_N\in\mathbb{N}_0:\sum_{i=1}^N p_i=p,\ \max_i p_i>1\right\}\right|.\end{aligned}$$

Within the above set, there are $p-1$ choices for $max_i p_i$ and one needs to account for the multiplicity of the maximal exponent. This explains the first two equalities in the following:

$$\begin{aligned}&\left|\left\{p_1,\dots,p_N\in\mathbb{N}_0:\sum_{i=1}^N p_i=p,\ \max_i p_i>1\right\}\right|\\ &=\sum_{j=2}^p\left|\left\{p_1,\dots,p_N\in\mathbb{N}_0:\sum_{i=1}^N p_i=p,\ \max_i p_i=j\right\}\right|\\ &=\sum_{j=2}^p\sum_{k=1}^{\lfloor p/j\rfloor}\left|\left\{p_1,\dots,p_N\in\mathbb{N}_0:\sum_{i=1}^N p_i=p,\ \max_i p_i=j,|\{i:p_i=j\}|=k\right\}\right|\\ &\leq\sum_{j=2}^p\sum_{k=1}^{\lfloor p/j\rfloor}N^k\left|\left\{p_1,\dots,p_{N-k}\in\mathbb{N}_0:\sum_{i=1}^{N-k} p_i=p-kj,\ \max_i p_i<j\right\}\right|\\ &\leq\sum_{j=2}^p\sum_{k=1}^{\lfloor p/j\rfloor}N^k\left|\left\{p_1,\dots,p_{N-k}\in\mathbb{N}_0:\sum_{i=1}^{N-k} p_i=p-kj\right\}\right|\\ &=\sum_{j=2}^p\sum_{k=1}^{\lfloor p/j\rfloor}N^k\binom{p-kj+N-k-1}{N-k-1}\\ &\leq\sum_{j=2}^p\sum_{k=1}^{\lfloor p/j\rfloor}N^{p-k(j-1)}\\ &\leq p^2N^{p-1}.\end{aligned}$$

The first inequality in the above follows from an over-estimate of the number of choices for the k indices with exponent $p_i = j$. The remaining inequalities are straightforward and the third equality follows from counting the number of ways to partition $p-kj$ into the ordered sum of $N-k$ values in $\mathbb{N}_0$.

Therefore,

$$\begin{aligned}\left|\mathbb{E}_N\left[\sum_{p_1+\dots+p_N=p,\max_i p_i>1}\binom{p}{p_1,\dots,p_N}\phi(X^{1,N})^{p_1}\cdots\phi(X^{N,N})^{p_N}\right]\right|&\leq p!||\phi||^pp^2N^{p-1}\\ &=o(N^p).\end{aligned}$$

The final equality in (3.2.2) follows from the fact that $p!\binom{N}{p}=N^p+o(N^p)$.

Sufficiency: By definition, see [56], $\frac{1}{N}\sum_{i=1}^{N}\delta_{X^{i,N}} \to \mu$ in distribution is equivalent to that $\langle \mu^N, \phi\rangle \xrightarrow{w} \langle \mu, \phi\rangle$ for all $\phi \in C_b$. For $f_i \in C_b$, $i = 1, \dots k$, write

$$\begin{aligned}
&\left|\mathbb{E}_N[f_1(X^{1,N})\cdots f_k(X^{k,N})] - \mathbb{E}\left[\prod_{i=1}^{k}\langle \mu, f_i\rangle\right]\right| \\
&\leq \left|\mathbb{E}_N\left[f_1(X^{1,N})\cdots f_k(X^{k,N}) - \prod_{i=1}^{k}\langle \mu^N, f_i\rangle\right]\right| + \left|\mathbb{E}_N\left[\prod_{i=1}^{k}\langle \mu^N, f_i\rangle\right] - \mathbb{E}\left[\prod_{i=1}^{k}\langle \mu, f_i\rangle\right]\right|.
\end{aligned} \tag{3.2.3}$$

The second term on the right hand side of (3.2.3) converges to zero as $N \to \infty$ since $f(\nu) := \prod_{i=1}^{k}\langle \nu, f_i\rangle$ is a continuous function with respect to the topology of weak convergence. The first term of the right hand side of (3.2.3) follows the argument provided by Sznitman [93]: The exchangeability allows one to rewrite

$$\begin{aligned}
&\left|\mathbb{E}_N\left[f_1(X^{1,N})\cdots f_k(X^{k,N}) - \prod_{i=1}^{k}\langle \mu^N, f_i\rangle\right]\right| \\
&\qquad = \left|\mathbb{E}_N\left[\frac{1}{N!}\sum_{\sigma\in S_N} f_1(X^{\sigma(1),N})\cdots f_k(X^{\sigma(k),N}) - \prod_{i=1}^{k}\langle \mu^N, f_i\rangle\right]\right| \\
&\qquad \leq \prod_{j=1}^{k}||f_j||\left(\left(\frac{(N-k)!}{N!} - \frac{1}{N^k}\right)\frac{N!}{(N-k)!} + \frac{1}{N^k}\left(N^k - \frac{N!}{(N-k)!}\right)\right) \\
&\qquad = 2\prod_{j=1}^{k}||f_j||\left(1 - \frac{N!}{N^k(N-k)!}\right) \\
&\qquad \to 0 \quad \text{as } N \to \infty.
\end{aligned} \tag{3.2.4}$$

The estimate follows from the fact that there are $N!/(N-k)!$ injections from $\{1, \dots, k\}$ into $\{1, \dots, N\}$, each appearing $(N-k)!$ times in the sum over the set of permutations, but only once in the expansion of the product $\prod_{i=1}^{k}\langle \mu^N, f_i\rangle$ and the fact that there are $N^k - N!/(N-k)!$ terms in the expansion of $\prod_{i=1}^{k}\langle \mu^N, f_i\rangle$ where there are repeated superscripts. The proof is complete. □

3.3 Conditional Propagation of Chaos

For this section, assume that solutions to the McKean-Vlasov with common noise (MKVCN) and the mean field particle system (MFSCN) exist. The forthcoming definition of conditional propagation of chaos is independent of the probability spaces on which the solutions are realised as it is a statement regarding the distributions of solutions. Nonetheless, it is convenient to introduce a particular coupling of the particle systems and conditionally independent copies of the solutions to (MKVCN).

To this end, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space supporting random elements $B, \mu, \{W^i, X^{i,\infty}\}_{i\in\mathbb{N}}, \{X^{i,N}, W^{i,N}\}_{i\in\mathbb{N}}$ and $\{B^N\}_{N\in\mathbb{N}}$ such that:

1. for each N, $(\{X^{i,N}\}_{i=1}^N, \{W^{i,N}\}_{i=1}^N, B^N)$ solves (MFSCN),
2. for each $i \in \mathbb{N}$, $(X^{i,\infty}, \mu, B, W^i)$ provides a weak solution to (MKVCN),
3. $X^{i,\infty}$ are all conditionally independent given (B, μ).

Definition 3.3.1 (Conditional Propagation of Chaos)**.** Say that conditional propagation of chaos occurs if and only if for any fixed $k \in \mathbb{N}$,

$$\text{Law}(X^{1,N}, ..., X^{k,N}) \to \text{Law}(X^{1,\infty}, ..., X^{k,\infty})$$

in the topology of weak convergence as $N \to \infty$. $\text{Law}(X^{1,\infty}, ..., X^{k,\infty})$ is the induced law under $\mathbb{P}$ of k conditionally independent copies of a solution to (MKVCN). Equivalently, for continuous and bounded f_i,

$$\mathbb{E}[f_1(X^{1,N}) \cdots f_k(X^{k,N})] \to \mathbb{E}[f_1(X^{1,\infty}) \cdots f_k(X^{k,\infty})].$$

Notice that in this case, the sequence of induced distributions of the particle systems are $\mathscr{L}(\mu)$-chaotic. Therefore, inspired by the trilogy of arguments given in Remark 1.5.3, and due to Proposition 3.2.2, the following heuristic is presented for demonstrating the conditional propagation of chaos.

1. Prove the weak existence, uniqueness and exchangeability of solutions to the particle systems $(X^{1,N}, ..., X^{N,N})_{N\in\mathbb{N}}$.
2. Establish tightness of the sequence of laws of μ^N in $\mathcal{P}(\mathcal{P}(\mathcal{C}))$ or equivalently the tightness of the laws of $\{X^{1,N}\}_{N\in\mathbb{N}}$ in $\mathcal{P}(\mathcal{C})$.
3. Prove that sub-sequential weak limits of $\mathscr{L}(X^{1,N}, \mu^N, W^{1,N}, B^N)$ are supported on the set of solutions to (MKVCN).
4. Establish the uniqueness of solutions to (MKVCN).

3.4 A Case Study

In this section, the above procedure is carried out for a combination of assumptions. Steps 1.-3. are verified for the simple case of boundedness, continuity and positive definiteness. Step 4. is established in a more complicated setting within Chapter 4.

Assumption 3.4.1. Let the following conditions hold:

1. (b, σ, ρ) are bounded and jointly continuous.
2. $\sigma\sigma^T + \rho\rho^T$ is everywhere positive definite.

In the following proposition, Assumption 4.3.1 is only used for the weak uniqueness of the weak solution to the McKean-Vlasov SDE with common noise.

Proposition 3.4.2. *Let Assumptions 3.4.1 and 4.3.1 hold. Then for any $N \in \mathbb{N}$, there exists a unique exchangeable family of weak solutions to (MFSCN) and a unique weak solution to the McKean-Vlasov SDE with common noise (MKVCN). Furthermore, there is conditional propagation of chaos.*

Proof. The proof will be completed in four stages:

1. Existence and uniqueness of weak solutions to (MFSCN) follows from [92]. Exchangeability follows from Assumption 3.4.1 and Theorem 2.2. in [98].

2. The tightness of $\{\mathscr{L}(\mu^N)\}_{N\in\mathbb{N}}$ or equivalently (by Proposition 1.5.2), the tightness of $\{\mathscr{L}(X^{1,N})\}_{N\in\mathbb{N}}$ follows from the boundedness of the coefficients.

3. Proving that any weakly convergent subsequence of $\{\mathscr{L}(X^{1,N},\mu^N,W^{1,N},B^N)\}$ converges to a solution to (MKVCN) requires a little more work. Some of the properties of a weak solution to the McKean-Vlasov with common noise are preserved under weak limits and are satisfied by the tuple $(X^{1,N},\mu^N,W^{1,N},B^N)$. Namely, the second compatibility condition and the first equation in (MKVCN) are preserved under weak limits in this case, this will be seen in the proof of Theorem 4.2.5. It remains to prove the second equation in (MKVCN) along with the first compatibility condition and the independence $\sigma(W_r - W_s : s \leq r \leq t) \perp\!\!\!\perp \mathcal{F}_t^{B,\mu} \vee \mathcal{F}_s^X$.

The independence follows from a slight adjustment of an argument given in [68] (Lemma 5.5). Let f, g, h be continuous bounded functions on the following spaces: f is defined on the state space of the private Brownian motions, $\mathbb{R}^{d_W}$, g is defined on the path space $\mathcal{C}$ and h is defined on the product space $C(\mathcal{I};\mathbb{R}^{d_B})\times\mathcal{P}(\mathcal{C})$. Introduce a coupling similar to the one defined prior to Definition 3.3.1, such that the particle systems are all supported on the same probability space along with processes X,μ,B,W that have the distribution of the sub-sequential weak limit. Abusing notation, let N denote the subsequence.

$$
\begin{aligned}
&\mathbb{E}[f(W_r^{1,N}-W_s^{1,N})g(X^{1,N}_{\cdot\wedge s})h(B^N,\mu^N)] - \mathbb{E}[f(W_r-W_s)]\mathbb{E}[g(X_{\cdot\wedge s})h(B,\mu)]\\
=&\mathbb{E}[f(W_r^{1,N}-W_s^{1,N})g(X^{1,N}_{\cdot\wedge s})h(B^N,\mu^N)]\\
&\qquad\qquad\qquad\qquad -\frac{1}{N}\sum_{i=1}^N \mathbb{E}[f(W_r^{i,N}-W_s^{i,N})g(X^{1,N}_{\cdot\wedge s})h(B^N,\mu^N)]\\
&+\Bigg(\frac{1}{N}\sum_{i=1}^N \mathbb{E}[f(W_r^{i,N}-W_s^{i,N})g(X^{1,N}_{\cdot\wedge s})h(B^N,\mu^N)]\\
&\qquad\qquad\qquad\qquad -\mathbb{E}[f(W_r-W_s)]\mathbb{E}[g(X^{1,N}_{\cdot\wedge s})h(B^N,\mu^N)]\Bigg)\\
&+\Big(\mathbb{E}[f(W_r-W_s)]\mathbb{E}[g(X^{1,N}_{\cdot\wedge s})h(B^N,\mu^N)] - \mathbb{E}[f(W_r-W_s)]\mathbb{E}[g(X_{\cdot\wedge s})h(B,\mu)]\Big)
\end{aligned}
\tag{3.4.1}
$$

The first term in the above may be handled using the following argument:

$$\begin{aligned}
&|\mathbb{E}[f(W_r^{1,N}-W_s^{1,N})g(X_{\cdot\wedge s}^{1,N})h(B^N,\mu^N)]\\
&\qquad\qquad\qquad -\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[f(W_r^{i,N}-W_s^{i,N})g(X_{\cdot\wedge s}^{1,N})h(B^N,\mu^N)]|\\
\leq&||g||\cdot||h||\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[|f(W_r^{1,N}-W_s^{1,N})-f(W_r^{i,N}-W_s^{i,N})|]\\
\leq&2||g||\cdot||h||\mathbb{E}[|f(W_r^{1,N}-W_s^{1,N})-f(W_r-W_s)|]\\
=&2||g||\cdot||h||\tilde{\mathbb{E}}[|f(\tilde{W}_r^{1,N}-\tilde{W}_s^{1,N})-f(\tilde{W}_r-\tilde{W}_s)|]\underset{N\to\infty}{\to}0,
\end{aligned}\tag{3.4.2}$$

by introducing the probability space $(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{\mathbb{P}})$ due to Skorokhod's representation theorem on which $\tilde{W}^{1,N}$ converges almost surely to $\tilde{W}$.

The second summand in (3.4.1) may be estimated as,

$$\begin{aligned}
&\left|\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[f(W_r^{i,N}-W_s^{i,N})g(X_{\cdot\wedge s}^{1,N})h(B^N,\mu^N)]\right.\\
&\qquad\qquad\qquad\left.-\mathbb{E}[f(W_r-W_s)]\mathbb{E}[g(X_{\cdot\wedge s}^{1,N})h(B^N,\mu^N)]\right| \qquad (3.4.3)\\
\leq&||g||\cdot||h||\mathbb{E}\left[\left|\frac{1}{N}\sum_{i=1}^{N}f(W_r^{i,N}-W_s^{i,N})-\mathbb{E}[f(W_r-W_s)]\right|\right].
\end{aligned}$$

Since the Brownian motions $W^{i,N}$ are i.i.d., by the law of large numbers the right hand side in the above converges to zero. The final summand in (3.4.1) converges to zero by the assumed weak convergence.

To prove the outstanding elements, by Remark 3.1.2, it is enough to show that $\mu_t=\mathscr{L}(X_{\cdot\wedge t}|\mathcal{F}_\infty^{B,\mu})$. This follows from another argument found in [68]:

$$\begin{aligned}
\mathbb{E}[\langle\mu_t,g\rangle h(B,\mu)]=&\lim_{N\to\infty}[\langle\mu^N,g\rangle h(B^N,\mu^N)]\\
=&\lim_{N\to\infty}\mathbb{E}[g(X_{\cdot\wedge t}^{1,N})h(B^N,\mu^N)]\\
=&\mathbb{E}[g(X_{\cdot\wedge t})h(B,\mu)]
\end{aligned}\tag{3.4.4}$$

4. Uniqueness of weak solutions to (MKVCN) will be established by Theorem 4.3.3 in Chapter 4.

The conditional propagation of chaos follows from Proposition 3.2.2. □

Chapter 4

McKean-Vlasov Dynamics with Common Noise: Existence and Uniqueness of Weak Solutions

This chapter demonstrates the existence and (joint) uniqueness of weak solutions to the McKean-Vlasov SDE with a common Brownian motion. Complementary to the previous chapter, justification of the definition of solutions used in this thesis is provided as a characterisation of weak limits of time-discretisation approximation schemes. The contents of this chapter and Appendix B are to appear in the Annals of Probability under the title 'Weak Existence and Uniqueness for McKean-Vlasov SDEs with Common Noise'.

4.1 Introduction and Literature Review

Throughout this chapter, let $I := \mathbb{R}^+$. Given a probability space supporting a random element Y and a sub-sigma algebra $\mathcal{G}$, let the regular conditional distribution of Y given $\mathcal{G}$, should it exist, be written $\mathscr{L}(Y|\mathcal{G})$. Henceforth, let X denote an $\mathbb{R}^{d_X}$-valued stochastic process and let μ denote a stochastic process valued on the space of probability measures on the path space of X. Additionally, ξ will be an $\mathbb{R}^{d_X}$-valued random vector and processes B and W are assumed to be Brownian motions of dimension d_B and d_W, respectively. The stochastic inputs B, W and ξ are assumed to be mutually independent. To establish the notation for this chapter, recall the McKean-Vlasov SDE with common noise:

$$\begin{aligned} X_t =& \xi + \int_0^t b(s, X_{\cdot\wedge s}, \mu_s)\, ds + \int_0^t \sigma(s, X_{\cdot\wedge s}, \mu_s)\, dW_s + \int_0^t \rho(s, X_{\cdot\wedge s}, \mu_s)\, dB_s, \\ \mu_s =& \mathscr{L}(X_{\cdot\wedge s}|\mathcal{F}_s^{B,\mu}). \end{aligned} \tag{MKVCN}$$

At first sight, the equation satisfied by the random measure flow μ seems strange, however, should μ be adapted to B, the measure flow satisfies $\mu_s = \mathscr{L}(X_{\cdot\wedge s}|\mathcal{F}_s^B)$ and (MKVCN) takes its more often seen form. Let $\mathcal{C}$ denote $C(I; \mathbb{R}^{d_X})$ equipped with the topology of uniform convergence on compact time intervals and $\mathcal{P}(\mathcal{C})$ denote the set of Borel probability measures on $\mathcal{C}$ equipped with the topology of weak convergence. Finally, let b, σ and ρ be measurable functions from $I \times \mathcal{C} \times \mathcal{P}(\mathcal{C})$ into $\mathbb{R}^{d_X}, \mathbb{R}^{d_X \times d_W}$ and $\mathbb{R}^{d_X \times d_B}$, respectively, that are always assumed to be at least progressive. Recall that a function f on $I \times \mathcal{C} \times \mathcal{P}(\mathcal{C})$ is called progressive if for any $t \in I$,

$$f(t, x, m) = f(t, x_{\cdot\wedge t}, m \circ \phi_t^{-1}), \text{ where } \phi_t : \mathcal{C} \ni x \mapsto x_{\cdot\wedge t} \in \mathcal{C}.$$

Under appropriate compatibility conditions and further specialisation of the coefficients b, σ and ρ it will be demonstrated that weak solutions to (MKVCN) yield measure valued solutions to the following SPDE that are both analytically and probabilistically weak. Analytically weak means that the solution is defined via its action on test functions and their derivatives. Probabilistically weak means that the measure valued solution process is not necessarily adapted to the stochastic input (a Brownian motion in this case). The SPDE solved is given as: $\mathbb{P}$-a.s. for all $t \in I$ and all $\varphi \in C_b^2(\mathbb{R}^{d_X})$

$$\langle \nu_t, \varphi \rangle = \langle \nu_0, \varphi \rangle + \int_0^t \langle \nu_s, L\varphi(s, \cdot, \nu_s) \rangle \, ds + \int_0^t \langle \nu_s, \partial_x \varphi \rho(s, \cdot, \nu_s) \rangle \, dB_s, \quad (4.1.1)$$

where $C_b^2(\mathbb{R}^{d_X})$ is the set of real valued functions on $\mathbb{R}^{d_X}$ with continuous and bounded mixed derivatives up to second order. Further, $\partial_x \varphi$ denotes the vector of first order derivatives of φ with respect to the components of x and the operator L acts on $C_b^2(\mathbb{R}^{d_X})$ test functions as follows:

$$L\varphi(t, x, \mu) := b(t, x, \mu)\partial_x \varphi + \frac{1}{2}\mathrm{tr}((\sigma\sigma^T + \rho\rho^T)(t, x, \mu)\partial_x^2 \varphi),$$

where $\partial_x^2 \varphi$ is the matrix of mixed second order derivatives with respect the components of x.

First Key Result: See Theorem 4.1.8 Assume that the coefficients b, σ and ρ are bounded and Markovian in the sense that $(b, \sigma, \rho)(t, x, m) = (b, \sigma, \rho)(t, x_t, m \circ \psi_t^{-1})$ where $\psi_t : \mathcal{C} \ni x \to x_t \in \mathbb{R}^{d_X}$. Then, the existence of a weak solution (to be defined) to the McKean-Vlasov SDE with common noise implies the existence of a measure valued solution the SPDE (4.1.1).

Motivated by the weak formulation of mean field games with common noise given by Carmona, Delarue and Lacker in [24], careful definitions of strong and weak solutions are given that facilitate this correspondence. In this framework, the statements can be brought in line with the generalisation of the well known equivalence of Yamada-Watanabe given by Kurtz in [65], justifying the form of the solution definitions. Secondly, this framework enables one to keep track of the dependence structure of approximations. This is key in allowing the use of compactness techniques, which are core to the weak existence result for the McKean-Vlasov SDE with common noise given in this chapter:

Second Key Result: See Theorem 4.2.5 There exists a weak solution to (MKVCN) of the type given in Definition 4.1.3 under assumptions of boundedness and joint continuity of the coefficients and integrability of the initial vector ξ.

The above theorem can be used to help establish an existence result for a particular class of coefficients:

Third Key Result: See Theorem 4.2.7 Assuming integrability of the initial condition and that the coefficients are Markovian, satisfy a non-degeneracy condition and their dependence on measure is of a linear integrated form with bounded measurable interaction kernel, the corresponding McKean-Vlasov SDE with common noise has a weak solution.

Strong uniqueness of solutions to the McKean-Vlasov SDE with common noise has been long established under the conditions of monotonicity [33] or Lipschitz continuity [67]. The final and main contribution of this chapter is to shed light on the question of uniqueness when the regularity of the coefficients is relaxed. In a non-degenerate setting, uniqueness in joint law for solutions to the McKean-Vlasov with common noise may be established:

Fourth Key Result: See Theorem 4.3.3 Assume that the diffusion coefficients σ and ρ do not depend upon measure and there exists a unique strong solution to the drift-less equation. Let the private noise coefficient σ satisfy a non-degeneracy condition and let $\sigma^{-1}b$ be total variation Lipschitz in the measure argument and bounded. Then, the equation (MKVCN) satisfies uniqueness in joint law.

The assumptions in the above result allow for only measurability (progressive) in the path argument of b with the price of non-degeneracy of the private noise coefficient σ. This extends a weak uniqueness argument employed in the case without common noise [20, 52, 70, 82, 84] to the case with a common noise. This idea of uniqueness proof, recently introduced by Mishura and Veretennikov [84], relies on representing two solutions by Girsanov Transformations from an intermediary probability space and estimating the total variation between the distribution of two solutions. Here, a particular Monge-Kantorovich problem for the path-distributions of solutions is studied, instead of the total variation distance, utilising a cost function tailored to this setting. It is easy to see that there is a non-empty intersection of the family of coefficients satisfying the assumptions of Theorem 4.2.7 and Theorem 4.3.3 for which joint weak existence-uniqueness holds.

Recently, there has been renewed interest in equations (MKVCN) and (4.1.1). A brief summary is presented below. This is roughly separated into two categories. The first category comprises of results related to McKean-Vlasov SDEs with common noise and/or stochastic partial differential equations (SPDEs) and the second includes those regarding mean field games with common noise.

Firstly, in contexts a little different from that of this chapter, Barbu, Röckner and Russo [6] consider a type of stochastic porous media equation and Briand et al. [18] study the problem of forwards and backwards SDEs where the distribution of any solution is constrained in some fashion and they extend their analysis to the common noise setting, where instead the conditional distributions are constrained. For well-posedness of a particular class of the McKean-Vlasov SDE with common noise and the corresponding SPDE, see the paper of Coghi and Gess [31] and see those of Kolokoltsov and Troeva [59, 62] for the sensitivity of solutions to perturbation of the initial data. For models motivated by application to finance and neuroscience, see Hambly and Søjmark [42] and Ledger and Søjmark [75]. Crisan, Janjigian and Kurtz [32] study a class of SPDEs that includes the Stochastic Allen-Cahn equation, extending the earlier work of Kurtz and Xiong [67] where strong solutions to an infinite system of mean field interacting particles driven by correlated noises are connected to strong solutions to a non-linear stochastic partial differential equation (SPDE) via the empirical distribution of the particles. Another approach to studying the types of SPDEs associated to particle systems driven by correlated noises is that of Dawson and Vaillancourt [33] who obtain measure-valued solutions of the aforementioned SPDE by studying the limit of empirical distributions to interacting systems of finitely many particles as the particle number increases to infinity.

In tandem, the mean field game theoretic framework introduced by Huang, Malhamé and Caines [47] and Lasry and Lions [74] has recently been subject to rapid development in the direction of common noise. For general theoretical results pertaining to well-posedness of the infinite player equilibrium and its closeness to the finite player equilibria, see [1, 24, 60, 61, 68] and the book of Cardaliaguet, Delarue, Lasry and Lions [21]. To see how the presence of a common noise can restore uniqueness to the mean field game, see the papers of Delarue and Tchuendom [34, 35, 96]. Cardaliaguet and Souganidis tackle the case without idiosyncratic noises in [23]. A substantial introduction to mean field games with common noise can be found in the second volume of the book of Carmona and Delarue [26]. The standard McKean-Vlasov setting with no common noise remains a popular field of study, with many new results. To list but a few: [5], [16], [44], [49], [50], [51], [30] and [88].

In summary, the key contributions of this chapter are as follows: first, an appropriate framework is developed which allows one to study weak solutions of McKean-Vlasov SDEs with common noise and, using the compatibility of solutions, connect them with weak solutions of SPDEs. Second, this framework allows the use of compactness arguments to obtain weak solutions to said equations and finally, a weak uniqueness result is obtained by a technique inspired by the method introduced in [84].

4.1.1 Definitions of Solutions

To begin, let $\mathbb{F}^{B,W,\xi} = \{\mathcal{F}_t^{B,W,\xi}\}_{t\in I}$ be defined by $\mathcal{F}_t^{B,W,\xi} := \mathcal{F}_t^B \vee \mathcal{F}_t^W \vee \sigma(\xi) = \sigma(B_s, W_s, \xi; 0 \le s \le t)$ for all $t \in I$ and similarly $\mathbb{F}^{B,\mu} = \{\mathcal{F}_t^{B,\mu}\}_{t\in I} := \{\mathcal{F}_t^B \vee \mathcal{F}_t^\mu\}_{t\in I} = \{\sigma(B_s, \mu_s; 0 \le s \le t)\}_{t\in I}$. When dealing with a measure space $(\Omega, \mathcal{F})$ equipped with multiple probability measures, say $\{\mathbb{P}^i\}_i$, denote the laws induced by a random element X under these measures as $\mathscr{L}^i(X)$. Vector and matrix norms will be denoted as $|\cdot|$ and L_p norms as $|\cdot|_{L_p}$. Consider the following definition of a strong solution to (MKVCN):

Definition 4.1.1 (Strong Solution to the McKean–Vlasov SDE with Common Noise). A filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ equipped with $\mathbb{F}$ Brownian motions B and W and initial condition ξ, all mutually independent, and an $\mathbb{F}$ adapted $\mathbb{R}^{d_X}$ valued process X is said to be a *strong solution* to the McKean-Vlasov SDE with common noise if the following conditions hold:

i) $\mathbb{P}$-a.s. for all $t \in I$, $\int_0^t (|b| + |\sigma|^2 + |\rho|^2)(s, X_{\cdot\wedge s}, \mathscr{L}(X_{\cdot\wedge s}|\mathcal{F}_s^B))\, ds < \infty$.

ii) X is $\mathbb{F}^{B,W,\xi}$ adapted.

iii) $\mathbb{P}$-a.s. for all $t \in I$,

$$X_t = \xi + \int_0^t b(s, X_{\cdot\wedge s}, \mathscr{L}(X_{\cdot\wedge s}|\mathcal{F}_s^B))\, ds + \int_0^t \sigma(s, X_{\cdot\wedge s}, \mathscr{L}(X_{\cdot\wedge s}|\mathcal{F}_s^B))\, dW_s + \int_0^t \rho(s, X_{\cdot\wedge s}, \mathscr{L}(X_{\cdot\wedge s}|\mathcal{F}_s^B))\, dB_s.$$

One can view a strong solution to the SDE (MKVCN) as a triple of stochastic inputs (B, W, ξ) defined on some probability space and a Borel measurable mapping $F : C(I;\mathbb{R}^{d_B}) \times C(I;\mathbb{R}^{d_W}) \times \mathbb{R}^{d_X} \to \mathbb{R}^{d_X}$ such that F maps the stochastic inputs (B, W, ξ) to an $\mathbb{F}^{B,W,\xi}$ adapted stochastic process $X := F(B, W, \xi)$ (the output) such that (X, B, W, ξ) satisfies (MKVCN). In the language of Kurtz [65] this is a strong compatible solution.

A guess at a good definition for a weak solution could be to remove the adaptedness requirement ii) from the above conditions and then ask that a weak solution should consist of a filtered probability space with the rest of Definition 4.1.1 unchanged. For clarity this is subsequently written (the choice of terminology 'weak-strong' will be justified after the definition).

Definition 4.1.2 (Weak-Strong Solution to the McKean–Vlasov SDE with Common Noise)**.** A weak-strong solution to (MKVCN) consists of a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ equipped with $\mathbb{F}$ Brownian motions B and W and initial condition ξ, all mutually independent, along with an $\mathbb{F}$ adapted $\mathbb{R}^{d_X}$ valued process X that satisfies the following conditions:

i) $\mathbb{P}$-a.s. for all $t \in I$, $\int_0^t (|b| + |\sigma|^2 + |\rho|^2)(s, X_{\cdot\wedge s}, \mathscr{L}(X_{\cdot\wedge s}|\mathcal{F}_s^B))\, ds < \infty$.

ii) $\mathbb{P}$-a.s. for all $t \in I$,

$$X_t = \xi + \int_0^t b(s, X_{\cdot\wedge s}, \mathscr{L}(X_{\cdot\wedge s}|\mathcal{F}_s^B))\, ds + \int_0^t \sigma(s, X_{\cdot\wedge s}, \mathscr{L}(X_{\cdot\wedge s}|\mathcal{F}_s^B))\, dW_s + \int_0^t \rho(s, X_{\cdot\wedge s}, \mathscr{L}(X_{\cdot\wedge s}|\mathcal{F}_s^B))\, dB_s.$$

There is an unfortunate shortcoming of such a definition. One can construct an example where weak solutions are expected to exist, but there are none of the above type. See counter-example 5.1 in [24] and Section ??. The issue is that one asks that the flow of conditional distributions μ from (MKVCN) should be adapted to the filtration generated by B and so whilst the process X might not be adapted to the stochastic inputs, the flow of conditional distributions must be. This justifies the terminology weak-strong. Since it is preferable to define weak solutions in such a way that they can be obtained under conditions comparable to the case without common noise, the relaxation to equation (MKVCN) will be made, justified by the following argument.

Since measurability is not generally preserved under weak limits, methods for approximating the flow of conditional distributions break down. To expand upon this point, imagine that one is solving a stochastic equation

$$\Gamma(Y, Z) = 0, \;\; Y \sim \nu.$$

The notation $Y \sim \nu$ means that the stochastic input Y has distribution ν. Z is the solution/output. Often, one seeks to solve the above by instead considering a mollified equation $\Gamma^n(Y, Z) = 0$, $Y \sim \nu$ such that "$\Gamma^n \to \Gamma$" and $\forall n$ the equation is *strongly* solvable; i.e. there is a measurable function F^n such that $Z^n := F^n(Y)$ is a solution. Then, passing to the limit in some sense "$\Gamma^n(Y, Z^n) \to \Gamma(Y, Z)$" one hopes to recover a solution to the original equation.

In the case of compactness arguments (weak existence), one may prove the weak convergence of a subsequence of the joint distributions of approximate solutions (Y, Z^n) and represent the solutions on a another probability space $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}})$ such that $(\bar{Y}^n, \bar{Z}^n) \to (\bar{Y}, \bar{Z})$ pointwise. Since $(\bar{Y}^n, \bar{Z}^n)$ have the same distribution as (Y, Z^n), one gets $F^n(\bar{Y}^n) = \bar{Z}^n$. Therefore $\bar{Z}$ is the pointwise limit of $\bar{Y}^n$ measurable functions, but unfortunately, $\bar{Y}^n$ varies along the same limit, and one cannot conclude that there is a measurable function F such that $\bar{Z} = F(\bar{Y})$. In fact, the existence of such a function corresponds to the existence of a strong solution.

The above observations give motivation to relax the measurability requirement of the regular conditional distribution appearing in the equation (MKVCN). Rather than asking that the measure argument of the coefficients be a version of $\mathscr{L}(X_{\cdot\wedge s}|\mathcal{F}^B_s)$, one should instead require that the argument be a flow of measures μ such that for any $s \in I$, $\mu_s = \mathscr{L}(X_{\cdot\wedge s}|\mathcal{F}^{B,\mu}_s)$. This relaxation is natural as, in general, this is the only way of identifying the limiting random measures obtained via weak convergence arguments.

Compatibility however, is preserved under weak limits when the marginal distribution of the stochastic inputs is fixed (see [7]). Due to this fact and the above motivation of connecting to the SPDE, a compatibility condition is introduced in the following definition.

Definition 4.1.3 (Weak Solution to the McKean–Vlasov SDE with Common Noise)**.** A weak solution to the McKean-Vlasov SDE with common noise consists of a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ equipped with $\mathbb{F}$ Brownian motions B and W and an $\mathcal{F}_0$ measurable random vector ξ, all mutually independent, along with $\mathbb{F}$ adapted processes X and μ that are $\mathbb{R}^{d_X}$ and $\mathcal{P}(\mathcal{C})$ valued respectively, satisfying the following conditions:

i) $\int_0^t (|b(s, X_{\cdot\wedge s}, \mu_s)| + |\sigma(s, X_{\cdot\wedge s}, \mu_s)|^2 + |\rho(s, X_{\cdot\wedge s}, \mu_s)|^2)ds < \infty$ $\mathbb{P}$-a.s. for all $t \in I$.

ii) X is compatible with (B, μ), (X, μ) is compatible with (B, W, ξ) and for $s, t \in I$ with $s \le t$, $\sigma(W_r - W_s : s \le r \le t) \perp\!\!\!\perp \mathcal{F}^{B,\mu}_t \vee \mathcal{F}^X_s$.

iii) $\mu_t = \mathscr{L}(X_{\cdot\wedge t}|\mathcal{F}^{B,\mu}_t)$ for all $t \in I$.

iv) $\mathbb{P}$-a.s. for all $t \in I$,

$$X_t = \xi + \int_0^t b(s, X_{\cdot\wedge s}, \mu_s)\, ds + \int_0^t \sigma(s, X_{\cdot\wedge s}, \mu_s)\, dW_s + \int_0^t \rho(s, X_{\cdot\wedge s}, \mu_s)\, dB_s. \tag{4.1.2}$$

In this definition, there is now a pair of outputs, (X, μ). As a weak solution, these outputs are allowed to have randomness external to that of the stochastic inputs, (ξ, B, W) (i.e. there is not a priori a Borel function G s.t. $(X, \mu) = G(B, W, \xi)$). Further, see that if condition ii) were removed, it would remain implied that (X, μ) is compatible with (B, W, ξ) since the processes B and W are assumed to be Brownian in the filtration $\mathbb{F}$ to which all processes are adapted and ξ is assumed $\mathcal{F}_0$ measurable. However, as these properties will need to be verified in the existence proof to prove that the limiting Brownian motions remain Brownian in the full filtration (generated by all limit processes), they are kept explicit in the definition.

To further justify considering the flow of measures μ as part of the solution pair, or 'stochastic outputs', note that it is desirable for the definition of a weak solution to be in accord with the Yamada-Watanabe principle.

Consider the solution as a pair (X, μ). Defining pathwise uniqueness such that for any two weak solutions (X, μ, B, W, ξ) and (X', μ', B, W, ξ) defined on the same probability space, (X, μ) and (X', μ') are indistinguishable. Then by way of the Yamada-Watanabe generalisation of Kurtz [65], assuming pathwise uniqueness, (X, μ) becomes $\mathbb{F}^{B,W,\xi}$ adapted and therefore, due to the independence structure, one can identify $\mu = \mathscr{L}(X|\mathcal{F}^B)$ and recover a strong solution of Definition 4.1.1. In keeping with the concept of a strong solution used by Kurtz in [65], the following simple proposition demonstrates that the notion of weak solution given by Definition 4.1.3 is appropriate.

Proposition 4.1.4. *A strong solution given by Definition 4.1.1 is equivalent to an $\mathbb{F}^{B,W,\xi}$ adapted weak solution pair (X, μ) of Definition 4.1.3.*

Proof. Given a strong solution of the type of Definition 4.1.1, (B, W, ξ, X), define a measure flow μ by $\mu_t := \mathscr{L}(X_{\cdot\wedge t}|\mathcal{F}_t^B)$. By definition, (X, μ, B, W, ξ) satisfies equation (4.1.2) and the integrability condition. Since μ is $\mathbb{F}^B$ adapted by construction, one has $\mathcal{F}_t^{B,\mu} = \mathcal{F}_t^B$ for all $t \in I$. Combining this fact with the $\mathbb{F}^{B,W,\xi}$ adaptedness of X, the conditions of Definition 4.1.3 are easily verified. For the converse direction, note that the independence of (W, ξ) and (B, μ) combined with the $\mathbb{F}^{B,W,\xi}$ adaptedness of μ implies that μ is $\mathbb{F}^B$ adapted. This in turn allows one to show that $\mu_t = \mathscr{L}(X_{\cdot\wedge t}|\mathcal{F}_t^{B,\mu}) = \mathscr{L}(X_{\cdot\wedge t}|\mathcal{F}_t^B)$ for all $t \in I$ and the equivalence follows. $\square$

Should one wish to obtain a weak solution via compactness arguments, when verifying the compatibility of X with (B, μ) for the weak limit, it becomes advantageous to work with $\mu_t := \mathscr{L}(X_{\cdot\wedge t}|\mathcal{F}_\infty^{B,\mu})$ and condition on the whole path of (B, μ). However, with the condition that X is compatible with (B, μ) in the sense that $\mathcal{F}_s^X$ is conditionally independent of $\mathcal{F}_t^{B,\mu}$ given $\mathcal{F}_s^{B,\mu}$ for any $s \le t \in I$, there is the following equivalence between characterisations of μ.

Proposition 4.1.5. *Given a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ equipped with continuous adapted processes X, B and μ, valued in $\mathbb{R}^{d_X}$, $\mathbb{R}^{d_B}$ and $\mathcal{P}(\mathcal{C})$ respectively, the following are equivalent:*

i) For all $t \in \mathcal{I}$, $\mu_t = \mathscr{L}(X_{\cdot\wedge t}|\mathcal{F}_t^{B,\mu})$ and X is compatible with (B, μ)

ii) For all $t \in \mathcal{I}$, $\mu_t = \mathscr{L}(X_{\cdot\wedge t}|\mathcal{F}_\infty^{B,\mu})$.

Remark 4.1.6. A consequence of either condition in the above proposition is that for all $s \in \mathcal{I}$ and any $t \in \mathcal{I} : s \leq t$, $\mu_s = \mathscr{L}(X_{\cdot\wedge s}|\mathcal{F}_t^{B,\mu})$. This property is proved in the beginning of the second half of the following proof.

Proof of Proposition 4.1.5. First it is shown that $i) \implies ii)$. Fix $t \in \mathcal{I}$ and let $f : \mathcal{C} \to \mathbb{R}$ and $g : C(\mathcal{I}; \mathbb{R}^{d_B}) \times C(\mathcal{I}; \mathcal{P}(\mathcal{C})) \to \mathbb{R}$ all be bounded and Borel measurable. Then,

$$\begin{aligned}
\mathbb{E}[f(X_{\cdot\wedge t})g(B, \mu)] &= \mathbb{E}[\mathbb{E}[f(X_{\cdot\wedge t})g(B, \mu)|\mathcal{F}_t^{B,\mu}]] \\
&= \mathbb{E}[\mathbb{E}[f(X_{\cdot\wedge t})|\mathcal{F}_t^{B,\mu}]\mathbb{E}[g(B, \mu)|\mathcal{F}_t^{B,\mu}]] \\
&= \mathbb{E}[\langle \mu_t, f\rangle \mathbb{E}[g(B, \mu)|\mathcal{F}_t^{B,\mu}]] \\
&= \mathbb{E}[\langle \mu_t, f\rangle g(B, \mu)].
\end{aligned}$$

The first equality follows from elementary properties of conditional expectation, the second from compatibility (see B.1.1 condition i), the third from definition of μ and the fourth from the measurability of the mapping $\mu_t \mapsto \langle \mu_t, f\rangle$ and hence the measurability of $\langle \mu_t, f\rangle$ with respect to the sigma algebra $\mathcal{F}_t^{B,\mu}$.

Since f and g are arbitrary bounded Borel measurable functions, the above equality holds for indicator functions $\mathbb{1}_F$ and $\mathbb{1}_G$ where $F \in \mathcal{B}(\mathcal{C})$ and $G \in \mathcal{B}(C(\mathcal{I}; \mathbb{R}^{d_B}) \times C(\mathcal{I}; \mathcal{P}(\mathcal{C})))$. Noting that μ_t is $\mathcal{F}_\infty^{B,\mu}$ measurable, μ_t satisfies the defining properties of the regular conditional distribution of $X_{\cdot\wedge t}$ given $\mathcal{F}_\infty^{B,\mu}$.

Now it remains to prove that $ii) \implies i)$. Using the fact that for $s, t \in \mathcal{I} : s \leq t$, μ_s is $\mathcal{F}_t^{B,\mu}$ measurable, and that for any $E \in \mathcal{F}_t^{B,\mu}$ and F defined as above, $\mathbb{E}[\mathbb{1}_F(X_{\cdot\wedge s})\mathbb{1}_E] = \mathbb{E}[\mu_s(F)\mathbb{1}_E]$ by definition of μ_s, μ_s can be identified as a version of the regular conditional distribution of $X_{\cdot\wedge s}$ given $\mathcal{F}_t^{B,\mu}$. I.e. for all $s, t \in \mathcal{I}$: $s \leq t$, $\mu_s = \mathscr{L}(X_{\cdot\wedge s}|\mathcal{F}_t^{B,\mu})$.

The first claim is immediate. To show compatibility, one needs to demonstrate the conditional independence of $\mathcal{F}_t^X$ from $\mathcal{F}_\infty^{B,\mu}$ given $\mathcal{F}_t^{B,\mu}$ (see again B.1.1 condition 1.). For fixed $t \in \mathcal{I}$, let f and g be as defined above and another function h be defined the same way as g. Then,

$$\begin{aligned}
&\mathbb{E}[\mathbb{E}[f(X_{\cdot\wedge t})g(B, \mu)|\mathcal{F}_t^{B,\mu}]h(B_{\cdot\wedge t}, \mu_{\cdot\wedge t})] \\
&= \mathbb{E}[\mathbb{E}[\mathbb{E}[f(X_{\cdot\wedge t})|\mathcal{F}_\infty^{B,\mu}]g(B, \mu)|\mathcal{F}_t^{B,\mu}]h(B_{\cdot\wedge t}, \mu_{\cdot\wedge t})] \\
&= \mathbb{E}[\mathbb{E}[\langle \mu_t, f\rangle g(B, \mu)|\mathcal{F}_t^{B,\mu}]h(B_{\cdot\wedge t}, \mu_{\cdot\wedge t})] \\
&= \mathbb{E}[\langle \mu_t, f\rangle \mathbb{E}[g(B, \mu)|\mathcal{F}_t^{B,\mu}]h(B_{\cdot\wedge t}, \mu_{\cdot\wedge t})] \\
&= \mathbb{E}[\mathbb{E}[f(X_{\cdot\wedge t})|\mathcal{F}_t^{B,\mu}]\mathbb{E}[g(B, \mu)|\mathcal{F}_t^{B,\mu}]h(B_{\cdot\wedge t}, \mu_{\cdot\wedge t})].
\end{aligned}$$

The first and third equalities follow from standard properties of conditional expectation and the second from the definition of μ. Finally, the fourth equality holds due to the observation at the beginning of this part of the proof. The conclusion holds by the uniqueness of conditional expectations. □

4.1.2 Associated SPDE

As mentioned in the introduction, assuming further structure of the coefficients, solutions to the McKean-Vlasov SDE with common noise correspond to measure valued solutions of a non-linear SPDE (4.1.1). The correspondence will be demonstrated in this subsection.

Definition 4.1.7 (Weak Solution to the SPDE (4.1.1))**.** A weak solution to the SPDE (4.1.1) is a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ equipped with an $\mathbb{F}$ Brownian motion B $\mathbb{F}$ adapted $\mathcal{P}(\mathbb{R}^{dx})$ valued process ν satisfying the equation (4.1.1), i.e.

$$\langle \nu_t, \varphi\rangle = \langle \nu_0, \varphi\rangle + \int_0^t \langle \nu_s, L\varphi(s, \cdot, \nu_s)\rangle\, ds + \int_0^t \langle \nu_s, \partial_x\varphi\rho(s, \cdot, \nu_s)\rangle\, dB_s$$

$\mathbb{P}$-a.s. for all $t \in I$ and for all test functions $\varphi \in C_b^2(\mathbb{R}^{dx})$.

Theorem 4.1.8. *Assume that the coefficients b, σ and ρ are bounded and* Markovian *in the sense that $(b, \sigma, \rho)(t, x, m) = (b, \sigma, \rho)(t, x_t, m \circ \psi_t^{-1})$ where $\psi_t : \mathcal{C} \ni x \to x_t \in \mathbb{R}^{dx}$. Then, the existence of a weak solution to the McKean-Vlasov SDE with common noise implies the existence of a weak solution the SPDE (4.1.1).*

Proof of Theorem 4.1.8. First, for any $\varphi \in C_0^\infty(\mathbb{R}^{dx})$, apply Itô's formula for $\varphi(X_t)$:

$$\begin{aligned}\varphi(X_t) =& \varphi(X_0) + \int_0^t L\varphi(s, X_s, \nu_s)\, ds \\ &+ \int_0^t \partial_x\varphi(X_s)\sigma(s, X_s, \nu_s)\, dW_s + \int_0^t \partial_x\varphi(X_s)\rho(s, X_s, \nu_s)\, dB_s\end{aligned}$$

where $\nu_s := \mu_s \circ \psi_s^{-1} = \mathscr{L}(X_s|\mathcal{F}_s^{B,\mu})$. Next, apply the conditional expectation with respect to $\mathcal{F}_t^{B,\mu}$ on both sides of the above equality:

$$\begin{aligned}\mathbb{E}[\varphi(X_t)|\mathcal{F}_t^{B,\mu}] =& \mathbb{E}[\varphi(X_0)|\mathcal{F}_t^{B,\mu}] + \mathbb{E}\left[\int_0^t L\varphi(s, X_s, \nu_s)\, ds\middle|\mathcal{F}_t^{B,\mu}\right] \\ &+ \mathbb{E}\left[\int_0^t \partial_x\varphi(X_s)\sigma(s, X_s, \nu_s)\, dW_s\middle|\mathcal{F}_t^{B,\mu}\right] \\ &+ \mathbb{E}\left[\int_0^t \partial_x\varphi(X_s))\rho(s, X_s, \nu_s)\, dB_s\middle|\mathcal{F}_t^{B,\mu}\right]\end{aligned}$$

Since φ has continuous compactly supported derivatives, and the coefficients b, σ, ρ are bounded, the integrands in the above expression are bounded and predictable. Therefore, one can apply the stochastic Fubini's theorem B.3.1 to the above stochastic integrals, identifying $\mathbb{F}^1$ as $\mathbb{F}^{B,\mu}$, $\mathbb{F}^2$ as $\mathbb{F}^{X,B,\mu}$, and $\mathbb{F}^3$ as $\mathbb{F}$.

$$\begin{aligned}&\langle \nu_t, \varphi\rangle \\ &= \langle \nu_0, \varphi\rangle + \int_0^t \mathbb{E}[L\varphi(s, X_s, \nu_s)|\mathcal{F}_s^{B,\mu}]\, ds + \int_0^t \mathbb{E}[\partial_x\varphi(X_s)\rho(s, X_s, \nu_s)|\mathcal{F}_s^{B,\mu}]\, dB_s \\ &= \langle \nu_0, \varphi\rangle + \int_0^t \langle \nu_s, L\varphi(s, \cdot, \nu_s)\rangle ds + \int_0^t \langle \nu_s, \partial_x\varphi\rho(s, \cdot, \nu_s)\rangle dB_s.\end{aligned}$$

□

Definition 4.1.9. A strong solution to the SPDE (4.1.1) is an $\mathbb{F}^B$-adapted weak solution.

Remark 4.1.10. If one can conclude that the flow of measures μ of a weak solution to the McKean-Vlasov SDE with common noise yields a strong solution to the SPDE, then one has a weak-strong solution of the type of Definition 4.1.2. This fact is exploited in [31], where Coghi and Gess establish a well-posedness result for (4.1.1).

4.2 Weak Existence

4.2.1 Assumptions

Assumption 4.2.1 (Coefficients)**.** Functions $b : I \times \mathcal{C} \times \mathcal{P}(\mathcal{C}) \to \mathbb{R}^d$, $\sigma : I \times \mathcal{C} \times \mathcal{P}(\mathcal{C}) \to \mathbb{R}^d \times \mathbb{R}^{d_W}$ and $\rho : I \times \mathcal{C} \times \mathcal{P}(\mathcal{C}) \to \mathbb{R}^{d_X} \times \mathbb{R}^{d_B}$ are *progressive* (i.e. for any $t \in I$, $(b, \sigma, \rho)(t, x, m) = (b, \sigma, \rho)(t, x_{\cdot\wedge t}, m \circ \phi_t^{-1})$, where $\phi_t : \mathcal{C} \ni x \mapsto x_{\cdot\wedge t} \in \mathcal{C}$), bounded and jointly continuous in the last two arguments in the following sense: if $(x_n \to x, m_n \overset{w}{\to} m)$ as $n \to \infty$ then $(b, \sigma, \rho)(t, x_n, m_n) \to (b, \sigma, \rho)(t, x, m)$ as $n \to \infty$.

Assumption 4.2.2 (Initial Condition)**.** For fixed $p \in (2, \infty]$, $|\xi|_{L_p} < \infty$.

Definition 4.2.3 (Euler-type Approximation Scheme)**.** Let $t_i^n := \frac{i}{n}$ for $i, n \in \mathbb{N}$, and define $\kappa_n(t) := t_i^n$ for $t \in [t_i^n, t_{i+1}^n)$. The sequence of Euler approximations X^n, are defined as strong solutions to the following distribution dependent SDEs constructed on a probability space supporting W, B and ξ. For all $n \in \mathbb{N}$, each X^n satisfies $\mathbb{P}$-a.s. for all $t \in I$,

$$\begin{aligned} X_t^n = \xi &+ \int_0^t b(s, X^n_{\cdot\wedge\kappa_n(s)}, \mathscr{L}(X^n_{\cdot\wedge\kappa_n(s)}|\mathcal{F}^B_{\kappa_n(s)}))\, ds \\ &+ \int_0^t \sigma(s, X^n_{\cdot\wedge\kappa_n(s)}, \mathscr{L}(X^n_{\cdot\wedge\kappa_n(s)}|\mathcal{F}^B_{\kappa_n(s)}))\, dW_s \\ &+ \int_0^t \rho(s, X^n_{\cdot\wedge\kappa_n(s)}, \mathscr{L}(X^n_{\cdot\wedge\kappa_n(s)}|\mathcal{F}^B_{\kappa_n(s)}))\, dB_s, \end{aligned} \tag{4.2.1}$$

Such solutions exist and can be constructed directly from the triple (ξ, B, W). Since for any $s \in I$ $X^n_{\cdot\wedge\kappa_n(s)}$ is $\mathcal{F}^{B,W,\xi}_{\kappa_n(s)}$ measurable,

$$\mathscr{L}(X^n_{\cdot\wedge\kappa_n(s)}|\mathcal{F}^B_{\kappa_n(s)}) = \mathscr{L}(X^n_{\cdot\wedge\kappa_n(s)}|\mathcal{F}^B_s) = \mathscr{L}(X^n_{\cdot\wedge\kappa_n(s)}|\mathcal{F}^B_\infty).$$

4.2.2 Auxiliary Lemmas

Lemma 4.2.4 (A Priori Estimates)**.** *Let Assumptions 4.2.1 and 4.2.2 hold. If $\{X^n\}_{n\in\mathbb{N}}$ is a (the) sequence of continuous stochastic processes satisfying* (4.2.1)*. Then for any $1 \le q \le p$ and $T < \infty$,*

$$\sup_n \mathbb{E}\left[\sup_{0\le t\le T} |X_t^n|^q\right] < \infty.$$

For any $q \geq 1$ and $s, t \in I$ such that $|t-s| \leq 1$,

$$\mathbb{E}\left[\sup_{s\leq u\leq t} |X^n_u - X^n_s|^q\right] \leq c_q(t-s)^{\frac{q}{2}}. \tag{4.2.2}$$

Proof. Is standard in the literature. See, for example, the proof of Theorem 21.9 in [55]. □

These estimate allow one to conclude tightness of the family $\{X^n\}_{n\in\mathbb{N}}$ by application of the Arzelà Ascoli characterisation of compact sets (see for example, problem 2.4.11 Karatzas and Shreve [57]) and prove that the family of flows of conditional measures constructed for the Euler approximations have continuous versions that induce a tight family of probability measures in $\mathcal{P}(C(I;\mathcal{P}_p(\mathcal{C})))$.

4.2.3 Existence Theorem

Theorem 4.2.5 (Existence of a Weak Solution to McKean-Vlasov SDE with Common Noise)**.** *Let Assumptions 4.2.1 and 4.2.2 hold. Then there exists a weak solution to the McKean-Vlasov SDE with common noise.*

Proof. There exists a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ satisfying the usual conditions, equipped with mutually independent $\mathbb{F}$ Brownian motions B and W and initial condition ξ. Construct the sequence of approximations X^n satisfying the Euler approximation SDE (4.2.1). This construction is carried out iteratively, applying Lemma B.3.2 on every interval of the approximation (of length $1/n$ for the n^{th} approximation) to ensure that the conditional distributions are valued in $\mathcal{P}_p(\mathcal{C})$. Note that the processes X^n are continuous by construction and are compatible with (B, W, ξ). It will now be demonstrated that the flow of measures $(\mathscr{L}(X^n_{\cdot\wedge t}|\mathcal{F}^B_t))_{t\geq 0}$ have continuous $\mathcal{P}_p(\mathcal{C})$ valued versions by verifying the conditions of Theorem B.2.1. The following holds for any $s, t \in I$ such that $|t-s| \leq 1$:

$$\begin{aligned}
&\mathbb{E}[W_p(\mathscr{L}(X^n_{\cdot\wedge t}|\mathcal{F}^B_t), \mathscr{L}(X^n_{\cdot\wedge s}|\mathcal{F}^B_s))^p] \\
&\qquad = \mathbb{E}[W_p(\mathscr{L}(X^n_{\cdot\wedge t}|\mathcal{F}^B_\infty), \mathscr{L}(X^n_{\cdot\wedge s}|\mathcal{F}^B_\infty))^p] \\
&\qquad \leq \mathbb{E}[\mathbb{E}[\sup_{s\leq u\leq t} |X^n_u - X^n_s|^p|\mathcal{F}^B_\infty]] \\
&\qquad \leq \mathbb{E}[\sup_{s\leq u\leq t} |X^n_u - X^n_s|^p] \\
&\qquad \leq c_{T,p}(t-s)^{\frac{p}{2}}
\end{aligned} \tag{4.2.3}$$

The equality follows from Proposition 4.1.5 and the inequalities follow consecutively from the definition of W_p, Jensen's inequality, properties of conditional expectation and Lemma 4.2.4. Since $p > 2$, there is a continuous modification (labelled μ^n) of each flow of measures via Theorem B.2.1. Moreover, by viewing ξ as the constant process $\{\Xi_t := \xi\}_{t\in I}$, see that $\mathscr{L}(X^n_{\cdot\wedge 0}|\mathcal{F}^B_0) = \mathscr{L}(X^n_{\cdot\wedge 0}) = \mathscr{L}(\Xi)$ is tight in $\mathcal{P}_p(\mathcal{C})$ as a Dirac mass and since the estimate (4.2.3) is uniform in n, the family of continuous modifications of the flows μ^n is tight in $C(I, \mathcal{P}_p(\mathbb{R}^{dx}))$ by application of Theorem B.2.2.

The family of joint distributions $\mathscr{L}((X^n, \mu^n, B, W)) =: \eta^n$ consequently defines a tight family of measures on $\mathcal{C} \times C(I; \mathcal{P}_p(\mathcal{C})) \times C(I; \mathbb{R}^{d_B}) \times C(I; \mathbb{R}^{d_W})$. By application of Prokhorov's theorem there is a subsequence $\{n_k\}_k$ and a probability measure η, such that $\eta^{n_k} \overset{w}{\to} \eta$.

Skorokhod's Representation theorem gives the existence of a probability space $(\tilde{\Omega}, \tilde{F}, \tilde{P})$ on which are defined random elements $\{\tilde{Z}^{n_k}\}_k$ and $\tilde{Z}$, valued on the above product space such that

$$\tilde{Z}^{n_k} \equiv (\tilde{X}^{n_k}, \tilde{\mu}^{n_k}, \tilde{B}^{n_k}, \tilde{W}^{n_k}) \sim \eta^{n_k}, \quad \tilde{Z} \equiv (\tilde{X}, \tilde{\mu}, \tilde{B}, \tilde{W}) \sim \eta$$

$$\text{and } \tilde{Z}^{n_k} \to \tilde{Z} \quad \tilde{\omega}\text{-surely}.$$

It is useful to note that independence/compatibility of one random element/process with respect to another is a property of the joint distribution. This fact will be used to verify a few properties of the constructed processes. Let the filtration $\tilde{\mathbb{F}}$ be defined as $\tilde{\mathcal{F}}_t := \sigma(\tilde{X}_s, \tilde{\mu}_s, \tilde{B}_s, \tilde{W}_s : s \le t)$. The adaptedness of the X and μ with respect to this filtration is immediate from the definition. That $\tilde{B}$ and $\tilde{W}$ are $\tilde{\mathbb{F}}$ Brownian motions will follow from the immersion of their natural filtrations in the filtration $\tilde{\mathbb{F}}$ and this will be verified later in the proof.

The proof will be concluded once the components of $\tilde{Z}$, $(\tilde{X}, \tilde{\mu}, \tilde{B}, \tilde{W})$ have be shown to satisfy items i) to iv) of Definition 4.1.3 with $\tilde{\xi} := \tilde{X}_0$. Item 1 follows from the boundedness of b, σ and ρ.

For the second item, it is easily checked that $\sigma(\tilde{W}_r - \tilde{W}_s : s \le r \le t) \perp\!\!\!\perp \mathcal{F}_t^{\tilde{B},\tilde{\mu}} \vee \mathcal{F}_s^{\tilde{X}}$ (see [8] Theorem 2.8). To show that $(\tilde{X}, \tilde{\mu})$ is compatible with $(\tilde{B}, \tilde{W}, \tilde{\xi})$, one needs to demonstrate the conditional independence of $\tilde{\mathcal{F}}_t^{\tilde{X},\tilde{\mu}}$ from $\tilde{\mathcal{F}}_\infty^{\tilde{B},\tilde{W},\tilde{\xi}}$ given $\tilde{\mathcal{F}}_t^{\tilde{B},\tilde{W},\tilde{\xi}}$. Let $f : C([0,t]; \mathbb{R}^{d_X} \times \mathcal{P}_p(\mathcal{C})) \to \mathbb{R}$ continuous and bounded, $g : C(I; \mathbb{R}^{d_B} \times \mathbb{R}^{d_W}) \times \mathbb{R}^{d_X} \to \mathbb{R}$ and $h : C([0,t]; \mathbb{R}^{d_B} \times \mathbb{R}^{d_W}) \times \mathbb{R}^{d_X} \to \mathbb{R}$ measurable and bounded. Let $X|_{[0,t]}$ denote the truncation of a process on I to its realisation on $[0,t]$. By application of Lemma 2.1 from [7],

$$\begin{aligned}
&\tilde{\mathbb{E}}[f((\tilde{X}, \tilde{\mu})|_{[0,t]})(g(\tilde{B}, \tilde{W}, \tilde{\xi}) - \tilde{\mathbb{E}}[g(\tilde{B}, \tilde{W}, \tilde{\xi})|\mathcal{F}_t^{\tilde{B},\tilde{W},\tilde{\xi}}])h((\tilde{B}, \tilde{W})|_{[0,t]}, \tilde{\xi})] \\
&= \lim_{k\to\infty} \tilde{\mathbb{E}}\Bigg[f((\tilde{X}^{n_k}, \tilde{\mu}^{n_k})|_{[0,t]}) \\
&\qquad \times \left(g(\tilde{B}^{n_k}, \tilde{W}^{n_k}, \tilde{\xi}^{n_k}) - \tilde{\mathbb{E}}[g(\tilde{B}^{n_k}, \tilde{W}^{n_k}, \tilde{\xi}^{n_k})|\mathcal{F}_t^{\tilde{B}^{n_k},\tilde{W}^{n_k},\tilde{\xi}^{n_k}}]\right) \\
&\qquad \times h((\tilde{B}^{n_k}, \tilde{W}^{n_k})|_{[0,t]}, \tilde{\xi}^{n_k})\Bigg] \\
&= \lim_{k\to\infty} \mathbb{E}\Bigg[f((X^{n_k}, \mu^{n_k})|_{[0,t]}) \\
&\qquad \times \left(g(B, W, \xi) - \mathbb{E}[g(B, W, \xi)|\mathcal{F}_t^{B,W,\xi}]\right) h((B, W)|_{[0,t]}, \xi)\Bigg] \\
&= 0.
\end{aligned}$$

The final equality holds since μ^{n_k} is a modification of a $\mathbb{F}^B$ adapted process on the space $(\Omega, \mathcal{F}, \mathbb{P})$ and X^{n_k} is a strong solution to the Euler scheme.

To see how to apply Lemma 2.1 from [7], notice that $\tilde{\mathbb{E}}[g(\tilde{B},\tilde{W},\tilde{\xi})|\mathcal{F}_t^{\tilde{B},\tilde{W},\tilde{\xi}}]$ is by definition $\mathcal{F}_t^{\tilde{B},\tilde{W},\tilde{\xi}}$ measurable and therefore by the Doob-Dynkin lemma (Lemma B.0.1) there exists a measurable function $G : C([0,t];\mathbb{R}^{d_B}\times\mathbb{R}^{d_W})\times\mathbb{R}^{d_X}\to\mathbb{R}$ such that $G((\tilde{B},\tilde{W})|_{[0,t]},\tilde{\xi}) = \tilde{\mathbb{E}}[g(\tilde{B},\tilde{W},\tilde{\xi})|\mathcal{F}_t^{\tilde{B},\tilde{W},\tilde{\xi}}]$. Since, $(\tilde{B},\tilde{W},\tilde{\xi})$ has the same distribution as $(\tilde{B}^{n_k},\tilde{W}^{n_k},\tilde{\xi}^{n_k})$,

$$\begin{aligned}
&\tilde{\mathbb{E}}[\tilde{\mathbb{E}}[g(\tilde{B}^{n_k},\tilde{W}^{n_k},\tilde{\xi}^{n_k})|\mathcal{F}_t^{\tilde{B}^{n_k},\tilde{W}^{n_k},\tilde{\xi}^{n_k}}]h((\tilde{B}^{n_k},\tilde{W}^{n_k})|_{[0,t]},\tilde{\xi}^{n_k})]\\
&\quad=\tilde{\mathbb{E}}[g(\tilde{B}^{n_k},\tilde{W}^{n_k},\tilde{\xi}^{n_k})h((\tilde{B}^{n_k},\tilde{W}^{n_k})|_{[0,t]},\tilde{\xi}^{n_k})]\\
&\quad=\tilde{\mathbb{E}}[g(\tilde{B},\tilde{W},\tilde{\xi})h((\tilde{B},\tilde{W})|_{[0,t]},\tilde{\xi})]\\
&\qquad\tilde{\mathbb{E}}[\tilde{\mathbb{E}}[g(\tilde{B},\tilde{W},\tilde{\xi})|\mathcal{F}_t^{\tilde{B},\tilde{W},\tilde{\xi}}]h((\tilde{B},\tilde{W})|_{[0,t]},\tilde{\xi})]\\
&\quad=\tilde{\mathbb{E}}[G((\tilde{B},\tilde{W})|_{[0,t]},\tilde{\xi})h((\tilde{B},\tilde{W})|_{[0,t]},\tilde{\xi})]\\
&\quad=\tilde{\mathbb{E}}[G((\tilde{B}^{n_k},\tilde{W}^{n_k})|_{[0,t]},\tilde{\xi}^{n_k})h((\tilde{B}^{n_k},\tilde{W}^{n_k})|_{[0,t]},\tilde{\xi}^{n_k})].
\end{aligned}$$

Therefore, the bounded and measurable function G provides a version of the conditional expectation appearing above, and the Lemma 2.1 from [7] can be applied.

It will be verified that for all $t\in I$, $\tilde{\mu}_t = \mathscr{L}(\tilde{X}_t|\mathcal{F}_\infty^{\tilde{B},\tilde{\mu}})$. Then, via Proposition 4.1.5, it holds that $\tilde{\mu}_t = \mathscr{L}(\tilde{X}_t|\tilde{\mathcal{F}}_t^{\tilde{B},\tilde{\mu}})$ for any $t\in I$ and $\tilde{X}$ is compatible with $(\tilde{B},\tilde{\mu})$. This verifies item iii) and the outstanding element of item ii). First, note that since $\tilde{\mu}$ is adapted to $\tilde{\mathbb{F}}^{\tilde{B},\tilde{\mu}}$ (the natural filtration of the tuple $\tilde{B},\tilde{\mu}$), all that needs to be verified to show that $\tilde{\mu}_t = \mathscr{L}(\tilde{X}_t|\tilde{\mathcal{F}}_\infty^{\tilde{B},\tilde{\mu}})$ for any $t\in I$ is that for $f:\mathcal{C}\to\mathbb{R}$ and $g: C(I;\mathbb{R}^{d_B})\times C(I;\mathcal{P}(\mathcal{C}))\to\mathbb{R}$ continuous and bounded,

$$\tilde{\mathbb{E}}[f(\tilde{X}_{\cdot\wedge t})g(\tilde{B},\tilde{\mu})] = \tilde{\mathbb{E}}[\langle\tilde{\mu}_t,f\rangle g(\tilde{B},\tilde{\mu})].$$

It will hold for f and g bounded and measurable by a Lusin's theorem approximation. The above equation holds since,

$$\begin{aligned}
\tilde{\mathbb{E}}[f(\tilde{X}_{\cdot\wedge t})g(\tilde{B},\tilde{\mu})] &= \lim_{k\to\infty}\tilde{\mathbb{E}}[f(\tilde{X}^{n_k}_{\cdot\wedge t})g(\tilde{B}^{n_k},\tilde{\mu}^{n_k})]\\
&= \lim_{k\to\infty}\mathbb{E}[f(X^{n_k}_{\cdot\wedge t})g(B,\mu^{n_k})]\\
&= \lim_{k\to\infty}\mathbb{E}[f(X^{n_k}_{\cdot\wedge t})g(B,\mathscr{L}(X^{n_k}|\mathcal{F}_\infty^B))]\\
&= \lim_{k\to\infty}\mathbb{E}[\mathscr{L}(X^{n_k}_{\cdot\wedge t}|\mathcal{F}_\infty^B)(f)g(B,\mathscr{L}(X^{n_k}|\mathcal{F}_\infty^B))] \qquad (4.2.4)\\
&= \lim_{k\to\infty}\mathbb{E}[\langle\mu_t^{n_k},f\rangle g(B,\mu^{n_k})]\\
&= \lim_{k\to\infty}\tilde{\mathbb{E}}[\langle\tilde{\mu}_t^{n_k},f\rangle g(\tilde{B}^{n_k},\tilde{\mu}^{n_k})]\\
&=\tilde{\mathbb{E}}[\langle\tilde{\mu}_t,f\rangle g(\tilde{B},\tilde{\mu})].
\end{aligned}$$

The first and last equalities follow from dominated convergence, the second and sixth from the fact that the joint distribution of (X^{n_k},B,μ^{n_k}) is the same as that of $(\tilde{X}^{n_k},\tilde{B}^{n_k},\tilde{\mu}^{n_k})$, the third and fifth equalities follow from the fact that $\{\mu_t^{n_k}\}_{t\in I}$ is a modification of $\{\mathscr{L}(X_t^{n_k}|\mathcal{F}_t^B)\}_{t\in I}$ and the compatibility of X^{n_k} with B, the fourth from the tower property of conditional expectation and definition of regular conditional distributions and the adaptedness of $\{\mathscr{L}(X_t^{n_k}|\mathcal{F}_t^B)\}_{t\in I}$ to $\mathbb{F}^B$. The convergence of

$\langle \tilde{\mu}_t^{n_k}, f\rangle$ to $\langle \tilde{\mu}_t, f\rangle$ follows from the fact that $\tilde{\mu}_t^{n_k} \to \tilde{\mu}_t$ $\tilde{\mathbb{P}}$-a.s. in $(\mathcal{P}_p(\mathcal{C}), W_p)$ - see Theorem 6.9 in [100].

Finally, the equation (4.1.2) will hold $\tilde{\mathbb{P}}$-a.s. for all $t \in I$ due to Lebesgue's dominated convergence theorem and a theorem due to Skorokhod (pg.32 [91]).

All items in the definition of a weak solution have been verified and thus the proof is concluded. □

4.2.4 Weak Existence for Bounded Measurable Interaction Kernel

Armed with Theorem 4.2.5, it is possible to prove the existence of weak solutions to a particular class of McKean-Vlasov SDEs with common noise, namely where the coefficients are bounded, measurable, non-degenerate, Markovian (in the sense that $(b,\sigma,\rho)(t,x,m) = (b,\sigma,\rho)(t,x_t,m\circ\psi_t^{-1})$ where $\psi_t : \mathcal{C} \ni x \to x_t \in \mathbb{R}^{d_X}$) and the dependence on measure is of the linear integrated form (this is sometimes referred to as a mean field interaction of scalar type). Hence, the spatial regularity of the coefficients can be relaxed at the price of a particular form of measure dependence. To be precise, the following assumption on the coefficients is formulated.

Assumption 4.2.6. The coefficients b, σ and ρ take the following form:

$$f(t,x,\nu) := \int \tilde{f}(t,x_t,y)\nu\circ\psi_t^{-1}(dy), \tag{4.2.5}$$

where f can be replaced with either b, σ or ρ. The functions (interaction kernels) $\tilde{b}$, $\tilde{\sigma}$ and $\tilde{\rho}$ are assumed to be bounded and measurable and, letting $\Sigma := (\sigma\,\rho)$,

$$\inf_{t,x,\nu}\ \inf_{\lambda\in\mathbb{R}^{d_X}:|\lambda|=1} \lambda^T\Sigma\Sigma^T\lambda > 0. \tag{4.2.6}$$

Theorem 4.2.7 (Weak Existence for Bounded Measurable Interaction Kernel)**.** *Under Assumption 4.2.6, the corresponding McKean-Vlasov SDE with common noise has a weak solution.*

Proof Outline. Similar to the proof of Mishura and Veretennikov [84] in the case without common noise, here the argument relies on a mollification of the interaction kernels $\tilde{b}$, $\tilde{\sigma}$ and $\tilde{\rho}$. The resulting mollified McKean-Vlasov SDEs with common noise have weak solutions by application of Theorem 4.2.5 and the solution processes satisfy the estimates given in Lemma 4.2.4. Therefore, a weakly convergent subsequence can be extracted from the sequence of joint laws of the approximate solutions. On a probability space given by the Skorokhod Representation Theorem, the limit process can be shown to be a solution to the original, un-mollified McKean-Vlasov SDE with common noise via application of estimates due to Krylov [64] (Ch.2 Sec.3 Thm.4).

Proof. First, the coefficients are mollified by replacing the interaction kernels with kernels $\tilde{b}^n$, $\tilde{\sigma}^n$ and $\tilde{\rho}^n$ that are defined by,

$$\tilde{f}^n(t,x,y) := n^{2d_X}\zeta(nx,ny) * \tilde{f}(t,x,y),$$

where ζ is a non-negative smooth function, vanishing for $|x|+|y|>1$, with $\int \zeta(x,y)dxdy = 1$. It is easy to see that the mollified coefficients satisfy the conditions of Theorem 4.2.5 and hence there exist weak solutions (X^n, μ^n, B^n, W^n) to the McKean-Vlasov SDEs with common noise defined by the mollified coefficients. Since the kernels' bounds are preserved by the mollification, the coefficients of the mollified McKean-Vlasov SDEs with common noise are uniformly bounded and therefore, by a standard procedure, the conclusion of Lemma 4.2.4 holds for this sequence of weak solutions. By the same argument from the proof of Theorem 4.2.5, one can extract a weakly convergent subsequence of the laws of these solutions. It will be convenient however, to consider another sequence of probability measures that gives access to copies of the solutions that are conditionally independent given (μ^n, B^n).

Denote the laws of the solutions (with ξ^i hidden inside X^i since $\xi^i = X_0^i$) by $\mathscr{L}(X^n, \mu^n, B^n, W^n)$. Disintegrate these distributions (see Chapter 10 in volume II of [12]) into the joint distribution of (μ^n, B^n) and the conditional distribution of (X^n, W^n) *given* μ^n, B^n. This is written as

$$\mathscr{L}(X^n, W^n, \mu^n, B^n)(dx, dw, d\nu, db) = p^n_{X,W}(dx, dw, \nu, b)\mathscr{L}(\mu^n, B^n)(d\nu, db).$$

Introducing a new sequence of probability distributions,

$$\pi^n(dx^1, dw^1, dx^2, dw^2, d\nu, db) := \prod_{i=1}^{2} p^n_{X,W}(dx^i, dw^i, \nu, b)\mathscr{L}(\mu^n, B^n)(d\nu, db)$$

and equipping the product space $\mathcal{C} \times C(\mathcal{I}; \mathbb{R}^{d_W}) \times \mathcal{C} \times C(\mathcal{I}; \mathbb{R}^{d_W}) \times C(\mathcal{I}; \mathcal{P}(\mathcal{C})) \times C(\mathcal{I}; \mathbb{R}^{d_B})$ with π^n, the canonical processes $(X, W, \hat{X}, \hat{W}, \mu, B)$ yields two weak solutions (X, W, μ, B) and $(\hat{X}, \hat{W}, \mu, B)$ with the property that (X, W) is conditionally independent of $(\hat{X}, \hat{W})$ given (μ, B). It is easy to see that the sequence π^n is also sequentially compact. As before, one extracts a weakly convergence subsequence and applies Skorokhod's Representation Theorem. Then, abusing notation to let n denote the subsequence, on some probability space there exists random elements $\{(X^n, W^n, \hat{X}^n, \hat{W}^n, \mu^n, B^n) \sim \pi^n\}_n$ and $(X, W, \hat{X}, \hat{W}, \mu, B) \sim \pi = \lim_n \pi^n$ such that $(X^n, W^n, \hat{X}^n, \hat{W}^n, \mu^n, B^n) \to (X, W, \hat{X}, \hat{W}, \mu, B)$ surely. The aim is to show that (X, W, μ, B) is a weak solution to the un-mollified McKean-Vlasov SDE with common noise. The first three items of Definition 4.1.3 are verified as in the proof of Theorem 4.2.5. The final item (that the SDE holds), however, requires additional consideration. It remains to show that

$$\int_0^t b^n(s, X^n, \mu^n)\, ds \to \int_0^t b(s, X, \mu)\, ds,$$
$$\int_0^t \sigma^n(s, X^n, \mu^n)\, dW^n_s \to \int_0^t \sigma(s, X, \mu)\, dW_s \text{ and}$$
$$\int_0^t \rho^n(s, X^n, \mu^n)\, dB^n_s \to \int_0^t \rho(s, X, \mu)\, dB_s$$

$\mathbb{P}$-a.s. for all $t \in \mathcal{I}$, again allowing n to denote the further subsequence taken to obtain this convergence. Consider some $t \in \mathcal{I} \cap \mathbb{Q}$, and the following sequence of estimates:

$$
\begin{aligned}
&\mathbb{E}\left[\left|\int_0^t b^n(s,X^n,\mu^n)ds-\int_0^t b(s,X,\mu)ds\right|\right]\\
\leq&\mathbb{E}\left[\int_0^t |b^n(s,X^n,\mu^n)ds-b(s,X,\mu)|ds\right]\\
\leq&\mathbb{E}\left[\int_0^t |b^n(s,X^n,\mu^n)-b^N(s,X^n,\mu^n)|ds\right]\\
&+\mathbb{E}\left[\int_0^t |b^N(s,X^n,\mu^n)-b^N(s,X,\mu)|ds\right]\\
&+\mathbb{E}\left[\int_0^t |b^N(s,X,\mu)-b(s,X,\mu)|ds\right]
\end{aligned}
$$

for some $N\in\mathbb{N}$. Then, by the form of the measure dependence of b and the Tower property,

$$
\begin{aligned}
&\mathbb{E}\left[\left|\int_0^t b^n(s,X^n,\mu^n)ds-\int_0^t b(s,X,\mu)ds\right|\right]\\
\leq&\mathbb{E}\left[\int_0^t\int |\tilde{b}^n-\tilde{b}^N|(s,X_s^n,y)\mu^n\circ\psi_s^{-1}(dy)ds\right]\\
&+\mathbb{E}\left[\int_0^t |b^N(s,X^n,\mu^n)-b^N(s,X,\mu)|ds\right]\\
&+\mathbb{E}\left[\int_0^t\int |\tilde{b}^N-\tilde{b}|(s,X_s,y)\mu\circ\psi_s^{-1}(dy)ds\right]\\
\leq&\int_0^t\mathbb{E}\left[\mathbb{E}\left[\int |\tilde{b}^n-\tilde{b}^N|(s,X_s^n,y)\mu^n\circ\psi_s^{-1}(dy)\middle|\mathcal{F}^{B^n,\mu^n}\right]\right]ds\\
&+\mathbb{E}\left[\int_0^t |b^N(s,X^n,\mu^n)ds-b^N(s,X,\mu)|ds\right]\\
&+\int_0^t\mathbb{E}\left[\mathbb{E}\left[\int |\tilde{b}^N-\tilde{b}|(s,X_s,y)\mu\circ\psi_s^{-1}(dy)\middle|\mathcal{F}^{B,\mu}\right]\right]ds.
\end{aligned}
\tag{4.2.7}
$$

The first term in the final line is handled as follows:

$$
\begin{aligned}
&\int_0^t\mathbb{E}\left[\mathbb{E}\left[\int |\tilde{b}^n-\tilde{b}^N|(s,X_s^n,y)\mu^n(dy)\middle|\mathcal{F}^{B^n,\mu^n}\right]\right]ds\\
=&\int_0^t\mathbb{E}\left[\iint |\tilde{b}^n-\tilde{b}^N|(s,x,y)\mu^n\circ\psi_s^{-1}(dx)\otimes\mu^n\circ\psi_s^{-1}(dy)\right]ds\\
=&\int_0^t\mathbb{E}[\mathbb{E}[|\tilde{b}^n-\tilde{b}^N|(s,X_s^n,\hat{X}_s^n)|\mathcal{F}^{B^n,\mu^n}]]ds\\
=&\int_0^t\mathbb{E}[|\tilde{b}^n-\tilde{b}^N|(s,X_s^n,\hat{X}_s^n)]ds\\
\leq&|\tilde{b}^n-\tilde{b}^N|_{L_{1+2d}}.
\end{aligned}
$$

The above equalities hold due to the construction of the measures π^n and the inequality by application of Theorem 4, Sec.3, Ch.2 of [64].

Repeating the above sequence of estimates with the superscript n removed, the final term of (4.2.7) can be dealt with leading to the estimate:

$$\mathbb{E}\Big[\Big|\int_0^t b^n(s,X^n,\mu^n)ds - \int_0^t b(s,X,\mu)ds\Big|\Big]$$
$$\leq |\tilde{b}^n - \tilde{b}^N|_{L_{1+2d}} + \mathbb{E}\Big[\int_0^t |b^N(s,X^n,\mu^n)ds - b^N(s,X,\mu)|ds\Big] + |\tilde{b}^N - \tilde{b}|_{L_{1+2d}}.$$

For any $\varepsilon > 0$, there is an N large enough such that for $n > N$, $|\tilde{b}^n - \tilde{b}^N|_{L_{1+2d}} + |\tilde{b}^N - \tilde{b}|_{L_{1+2d}} < \varepsilon/2$. Also, as $n \to \infty$, by the continuity of b^N, the middle term in the above inequality vanishes. Therefore, for each $N \in \mathbb{N}$, there is an n_N such that for all $n > n_N$, the middle term is bounded by $\varepsilon/2$ and therefore,

$$\int_0^t b^n(s,X^n,\mu^n)ds \xrightarrow{\mathbb{P}} \int_0^t b(s,X,\mu)ds$$

for any $t \in I \cap \mathbb{Q}$. This can be elevated to almost sure convergence along a subsequence and to all $t \in I$ by continuity. To prove the corresponding limits for the stochastic integrals, one follows an analogous procedure to that of the drift convergence. Writing f, M in place of σ, W or ρ, B, one can estimate as follows:

$$\begin{aligned}
&(1/3)\mathbb{E}\Big[\Big(\int_0^t f^n(s,X^n,\mu^n)dM_s^n - \int_0^t f(s,X,\mu)dM_s\Big)^2\Big]\\
\leq&\mathbb{E}\Big[\Big(\int_0^t (f^n(s,X^n,\mu^n) - f^N(s,X^n,\mu^n))dM_s^n\Big)^2\Big]\\
&+\mathbb{E}\Big[\Big(\int_0^t f^N(s,X^n,\mu^n)dM_s^n - \int_0^t f^N(s,X,\mu)dM_s\Big)^2\Big]\\
&+\mathbb{E}\Big[\Big(\int_0^t (f^N(s,X,\mu) - f(s,X,\mu))dM_s\Big)^2\Big]
\end{aligned} \tag{4.2.8}$$

for some $N \in \mathbb{N}$. To finish, apply the Itô isometry to the first and third terms on the right hand side of (4.2.8) and follow an almost exactly analogous procedure as with the drift convergence, taking care of the second power appearing. Handle the second term with Skorokhod's lemma for the convergence of stochastic integrals, see [91] pg.32. One arrives at the following estimate:

$$\begin{aligned}
&(1/3)\mathbb{E}\Big[\Big(\int_0^t f^n(s,X^n,\mu^n)dM_s^n - \int_0^t f(s,X,\mu)dM_s\Big)^2\Big]\\
\leq&|\tilde{f}^n - \tilde{f}^N|^2_{L_{2(1+2d)}} + \mathbb{E}\Big[\Big(\int_0^t f^N(s,X^n,\mu^n)dM_s^n - \int_0^t f^N(s,X,\mu)dM_s\Big)^2\Big]\\
&+|\tilde{f}^N - \tilde{f}|^2_{L_{2(1+2d)}}\\
<&\varepsilon
\end{aligned}$$

for sufficiently large n depending on the choice of $\varepsilon > 0$. □

4.3 Uniqueness in Joint Law

In this section, a particular class of equations of the type (4.1.2) will be studied. Namely, the case where the diffusion coefficients σ and ρ do not depend upon measure. The authors expect that with similar techniques to those given in [82] and [84] the result here can be extended to include some spatial growth. However, in the interest of conveying how one overcomes the barriers of extending this method to the common noise setting without becoming mired in additional technical difficulties, the following assumptions are made regarding the coefficients.

Assumption 4.3.1. The coefficients b, σ and ρ are measurable and progressive. The coefficients σ and ρ do not depend on the measure argument and are such that there exists a unique strong solution to the driftless SDE:

$$dX_t^0 = \sigma(t, X^0)dW_t + \rho(t, X^0)dB_t. \tag{4.3.1}$$

Further, $d_X = d_W$, σ is non-degenerate, invertible and $\sigma^{-1}b$ is bounded and Lipschitz continuous in the measure component with respect to the total variation distance, i.e. there is a constant c_{TV} such that

$$|\sigma(t,x)^{-1}b(t,x,\mu) - \sigma(t,x)^{-1}b(t,x,\nu)| \le c_{\mathrm{TV}} d_{\mathrm{TV}}(\mu,\nu).$$

Under the above assumption, the McKean-Vlasov SDE with common noise, (4.1.2), takes the form:

$$X_t = \xi + \int_0^t b(s, X_{\cdot\wedge s}, \mu_s)\, ds + \int_0^t \sigma(s, X_{\cdot\wedge s})\, dW_s + \int_0^t \rho(s, X_{\cdot\wedge s})\, dB_s. \tag{4.3.2}$$

Definition 4.3.2 (Uniqueness in Joint Law)**.** The McKean-Vlasov SDE with common noise is said to satisfy 'uniqueness in joint law' if any two weak solutions (in the sense of Definition 4.1.3), $(X^1, \mu^1, B^1, W^1, \xi^1)$ and $(X^2, \mu^2, B^2, W^2, \xi^2)$ have the same *joint* distribution.

Theorem 4.3.3. *Under Assumption 4.3.1, the McKean-Vlasov SDE with common noise of the form* (4.3.2) *satisfies uniqueness in joint law.*

The proof of Theorem 4.3.3 will be given in Subsection 4.3.2. The following subsection provides a lemma that establishes uniqueness in joint law for the SDEs with random coefficients obtained when one considers the measure valued process provided by a weak solution to (4.3.2) as a stochastic input.

4.3.1 Auxiliary Lemma

Definition 4.3.4. A filtered probability space supporting Brownian motions W and B, an adapted stochastic process μ and an $\mathcal{F}_0$ measurable random vector ξ, such that $(B,\mu) \perp\!\!\!\perp (W,\xi)$ is said to be a weak solution on $[0,T]$ to the SDE with random coefficients:

$$X_t = \xi + \int_0^t b(s, X, \mu)ds + \int_0^t \sigma(s, X)dW_s + \int_0^t \rho(s, X)dB_s, \tag{4.3.3}$$

if it also supports an adapted process X, such that

1. $\mathbb{P}$-a.s. $\forall\ t \in [0, T]$, $\int_0^t |b(s, X, \mu)| + |\sigma(s, X, \mu)|^2 + |\rho(s, X, \mu)|^2\, ds < \infty$.
2. X, μ, B, W, ξ satisfy (4.3.3) $\mathbb{P}$-a.s. $\forall\ t \in [0, T]$.

Lemma 4.3.5. *Under Assumption 4.3.1, the SDE with random coefficients (4.3.3) satisfies joint uniqueness in law on $[0, T]$ for any $T < \infty$.*
Which is to say that given any two weak solutions of type of Definition 4.3.4, $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1, X^1, \mu^1, B^1, W^1, \xi^1)$ and $(\Omega^2, \mathcal{F}^2, \mathbb{P}^2, X^2, \mu^2, B^2, W^2, \xi^2)$ such that $\mathscr{L}^1(\mu^1, B^1, W^1, \xi^1) = \mathscr{L}^2(\mu^2, B^2, W^2, \xi^2)$, the joint distributions of the solutions $\mathscr{L}^1(X^1_{\cdot\wedge T}, \mu^1, B^1, W^1, \xi^1)$ and $\mathscr{L}^2(X^2_{\cdot\wedge T}, \mu^2, B^2, W^2, \xi^2)$ are equal.

Proof. Given an arbitrary solution (X, μ, B, W, ξ) to (4.3.3) on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with $\nu := \mathscr{L}(\mu, B, W, \xi)$, define an equivalent probability measure $\mathbb{Q}_T$ by

$$\frac{d\mathbb{Q}_T}{d\mathbb{P}} := \mathcal{E}_T\Big(- \int_0^{\cdot} \sigma^{-1}(s, X) b(s, X, \mu) dW_s \Big).$$

As $(\mu, B, \xi) \perp\!\!\!\perp W$, the tuple (μ, B, ξ) has the same joint distribution under $\mathbb{Q}_T$ or $\mathbb{P}$. By Girsanov's Theorem, $\tilde{W} := W + \int_0^{\cdot\wedge T} \sigma^{-1}(s, X) b(s, X, \mu) ds$ is a $\mathbb{Q}_T$-Brownian motion. Therefore, $(\mu, B, \tilde{W}, \xi) \sim \nu$ under $\mathbb{Q}_T$. Also, since X satisfies (4.3.1) on $[0, T]$ under $\mathbb{Q}_T$, with stochastic input $(B, \tilde{W}, \xi)$, the process $X_{\cdot\wedge T}$ has a uniquely determined law on $\mathbb{Q}_T$ since (4.3.1) has a unique strong solution.

Combining these facts, under $\mathbb{Q}_T$, $(X_{\cdot\wedge T}, \mu, B, \tilde{W}, \xi)$ has a joint distribution that does not depend upon the choice of weak solution. This uniquely determines their joint law with W and $\mathcal{E}_T(\int_0^{\cdot} \sigma^{-1}(s, Y) b(s, Y, G(U, B)) d\tilde{W}_s)$ under $\mathbb{Q}_T$.

Since $\mathbb{P}$ and $\mathbb{Q}_T$ are equivalent,

$$\mathbb{P}[(X_{\cdot\wedge T}, \mu, B, W, \xi) \in A] = \mathbb{E}_{\mathbb{Q}_T}\left[\frac{d\mathbb{P}}{d\mathbb{Q}_T} \mathbb{1}_{(X_{\cdot\wedge T}, \mu, B, W, \xi) \in A} \right].$$

Further, since $\frac{d\mathbb{P}}{d\mathbb{Q}_T} = \left(\frac{d\mathbb{Q}_T}{d\mathbb{P}}\right)^{-1}$ one can write,

$$\begin{aligned}\frac{d\mathbb{P}}{d\mathbb{Q}_T} &= \exp\left\{ \int_0^T \sigma^{-1}(s, X) b(s, X, \mu) dW_s + \frac{1}{2}\int_0^T |\sigma^{-1}(s, X) b(s, X, \mu)|^2 ds \right\} \\ &= \exp\left\{ \int_0^T \sigma^{-1}(s, X) b(s, X, \mu) d\tilde{W}_s - \frac{1}{2}\int_0^T |\sigma^{-1}(s, X) b(s, X, \mu)|^2 ds \right\} \\ &= \mathcal{E}_T\Big(\int_0^{\cdot} \sigma^{-1}(s, X) b(s, X, \mu) d\tilde{W}_s \Big).\end{aligned} \tag{4.3.4}$$

Finally,

$$\begin{aligned}&\mathbb{P}[(X_{\cdot\wedge T}, \mu, B, W, \xi) \in A] \\ =&\mathbb{E}_{\mathbb{Q}_T}\left[\frac{d\mathbb{P}}{d\mathbb{Q}_T} \mathbb{1}_{(X_{\cdot\wedge T}, \mu, B, W, \xi) \in A} \right] \\ =&\mathbb{E}_{\mathbb{Q}_T}\left[\mathcal{E}_T\Big(\int_0^{\cdot} \sigma^{-1}(s, X) b(s, X, \mu) d\tilde{W}_s \Big) \mathbb{1}_{(X, \mu, B, \tilde{W} - \int_0^{\cdot\wedge T} \sigma^{-1}(s, X) b(s, X, \mu) ds, \xi) \in A} \right],\end{aligned}$$

which does not depend upon the choice of weak solution. □

4.3.2 Proof of the Uniqueness Theorem

To aid in the reading of this subsection, the strategy is briefly outlined as follows:

Proof Outline

Steps 1.-2. Disintegrate the joint distributions of the solutions to identify the underlying randomness behind the flows of conditional distributions (μ^1 and μ^2).

Steps 3.-4. Introduce a Monge-Kantorovich Problem with a tailored cost function that forces the optimal coupling for this problem to constrain the underlying randomness to be the same for each solution.

Step 5. Show that it is possible to represent the distributions of the solutions by a unique solution to the drift-less equation viewed on two probability spaces related by Girsanov transformations. This requires the uniqueness in law to a certain class of SDEs with random coefficients as given by Lemma 4.3.5.

Step 6. For a small time interval, estimate the distance between two processes' distributions by studying the dual Kantorovich Problem, showing that for a small time interval, there is uniqueness in joint law.

Step 7. Conclude by induction.

Proof of Theorem 4.3.3. Given two weak solutions to (4.3.2) of the form given by Definition (4.1.3),
$(X^1, \mu^1, B^1, W^1, \xi^1)$ and $(X^2, \mu^2, B^2, W^2, \xi^2)$, denote the laws of the solutions (with ξ^i hidden inside X^i since $\xi^i = X^i_0$) on their respective probability spaces by

$$\mathscr{L}^1(X^1, \mu^1, B^1, W^1) \text{ and } \mathscr{L}^2(X^2, \mu^2, B^2, W^2),$$

where the superscript on $\mathscr{L}$ refers to the fact that these weak solutions may be defined on different probability spaces. In order to compare the distributions of the two solutions, one needs to couple the distributions on a probability space in such a way that fixes the underlying randomness of both μ^1 and μ^2 to be the same. This is done as follows:

1. Disintegrate the joint distributions of the two solutions (see Chapter 10 in volume II of [12]) into the joint distributions of (μ^i, B^i, W^i) and the conditional distribution of X^i *given* μ^i, B^i, W^i. This is written as

$$\mathscr{L}^i(X^i, \mu^i, B^i, W^i) = p^i_X(dx, \mu, b, w)\mathscr{L}^i(\mu^i, B^i)(d\mu, db)\mathscr{L}^i(W^i)(dw),$$

using the independence of W^i and (μ^i, B^i).

2. From Blackwell and Dubins [9], there exists for each $i \in \{1, 2\}$, a measurable function $G^i : [0, 1] \times C(I; \mathbb{R}^{d_B}) \to C(I; \mathcal{P}(\mathcal{C}))$, such that, if on some probability

space there are elements U, B such that $U \sim \text{Unif}(0,1) =: \lambda$, $B \sim \mathscr{L}^i(B^i)$ and $U \perp\!\!\!\perp B$, then

$$\mathscr{L}(G^i(U,B),B) = \mathscr{L}^i(\mu^i, B^i).$$

Note that the functions G^i cannot be claimed to be *adapted* in the sense that, if for $b^1, b^2 \in C(I;\mathbb{R}^{d_B})$ such that $b^1_{\cdot\wedge t} = b^2_{\cdot\wedge t}$ for some $t \in I$, then $G^i(u,b^1)_t = G^i(u,b^2)_t$. This is shown in Example 7.3 of [7].

Letting $\mathcal{W}_d$ denote Wiener measure on $C(I;\mathbb{R}^d)$, consider for $i \in \{1,2\}$,

$$\pi^i := p^i_X(dx,\mu,b,w)\delta_{G^i(u,b)}(d\mu)\lambda(du)\mathcal{W}_{d_B}(db)\mathcal{W}_{d_W}(dw).$$

Equipping the space $E := (\mathcal{C} \times C(I;\mathcal{P}(\mathcal{C})) \times [0,1] \times C(I;\mathbb{R}^{d_B}) \times C(I;\mathbb{R}^{d_W}))$ and its product σ-algebra with the measure π^i, the canonical random elements (X,μ,U,B,W) are such that (X,μ,B,W) have distribution $\mathscr{L}^i(X^i,\mu^i,B^i,W^i)$.

Further, for $i \in \{1,2\}$, introduce the measure

$$\pi^i_X := p^i_X(dx, G^i(u,b), b, w)\lambda(du)\mathcal{W}_{d_B}(db)\mathcal{W}_{d_W}(dw).$$

One can equip the product space $E^* := (\mathcal{C} \times [0,1] \times C(I;\mathbb{R}^{d_B}) \times C(I;\mathbb{R}^{d_W}))$ (with product σ-algebra denoted $\mathcal{B}(E^*)$) with π^i_X and define $\mu := G^i(U,B)$. Then, the canonical random elements X,U,B,W along with μ satisfy again, $\mathscr{L}^{\pi^i_X}(X,\mu,B,W) = \mathscr{L}^i(X^i,\mu^i,B^i,W^i)$ and consequently, denoting $(\Omega,\mathcal{F},\mathbb{P}) := (E^*,\mathcal{B}(E^*),\pi^i_X)$, for any $A \in \mathcal{B}(\mathcal{C})$ and bounded measurable $f : C(I;\mathcal{P}(\mathcal{C})) \times C(I;\mathbb{R}^{d_B}) \to \mathbb{R}$,

$$\begin{aligned}\mathbb{E}[G^i(U,B)_t(A)f(G^i(U,B),B)] =& \mathbb{E}^i[\mu^i_t(A)f(\mu^i,B^i)] \\ =& \mathbb{E}^i[\mathbb{1}_A(X^i_{\cdot\wedge t})f(\mu^i,B^i)] \\ =& \mathbb{E}[\mathbb{1}_A(X_{\cdot\wedge t})f(G^i(U,B),B)].\end{aligned} \tag{4.3.5}$$

Hence, $\mu_t = G^i(U,B)_t = \mathscr{L}(X_{\cdot\wedge t}|G^i(U,B),B) = \mathscr{L}(X_{\cdot\wedge t}|\mu,B)$ for all $t \in I$. An important observation is that, since X is independent of U given $\sigma(G^i(U,B),B)$, $\mu_t = \mathscr{L}(X_{\cdot\wedge t}|U,B)$ for all $t \in I$.

3. On the product space $E^* \times E^*$, define the lower semi-continuous cost function

$$c^*((x^1,u^1,b^1,w^1),(x^2,u^2,b^2,w^2)) := \begin{cases} \mathbb{1}_{x^1\neq x^2} + d(w^1,w^2)\wedge 1 \\ \qquad \text{if } (u^1,b^1) = (u^2,b^2), \\ \infty \qquad \text{otherwise.}\end{cases} \tag{4.3.6}$$

where d is the uniform metric on $C(I;\mathbb{R}^{d_W})$. Let W^* be the Monge-Kantorovich Problem (see Chapters 4 and 5 in [100]) with cost function c^*:

$$W^*(\pi^1_X,\pi^2_X) := \inf_{\pi:\,\pi \text{ couples } \pi^1_X,\pi^2_X} \int_{E^*\times E^*} c^* d\pi. \tag{4.3.7}$$

There exists an optimal coupling for this problem (a coupling minimizing the expected cost $\int c^* d\pi$) since the cost function c^* is lower semi-continuous, see [100], Theorem 4.1. If $W^*(\pi^1_X,\pi^2_X) = 0$, then one can conclude $\pi^1_X = \pi^2_X$

since $c^*((x^1,u^1,b^1,w^1),(x^2,u^2,b^2,w^2)) = 0$ if and only if $(x^1,u^1,b^1,w^1) = (x^2,u^2,b^2,w^2)$. Further, on the optimal coupling from (4.3.7), following the argument behind equation (4.3.5),

$$G^1(U,B)_t = \mathscr{L}(X^1_{\cdot\wedge t}|U,B) = \mathscr{L}(X^2_{\cdot\wedge t}|U,B) = G^2(U,B)_t,$$

almost surely for all $t \in I$, which by the continuity of sample paths of $G^i(U,B)$ is enough to claim that $G^1(U,B)$ and $G^2(U,B)$ are almost surely equal. It will consequently be the aim to show $W^*(\pi^1_X,\pi^2_X) = 0$ for any two solutions to (4.3.2). First, note that by the gluing lemma there exists a probability space $(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{\mathbb{P}})$ on which there are random elements $\tilde{X}^1,\tilde{X}^2,\tilde{U},\tilde{B},\tilde{W}^1,\tilde{W}^2$ with $\tilde{\mathscr{L}}(\tilde{X}^i,\tilde{U},\tilde{B},\tilde{W}^i) = \pi^i_X$. It is easy to see that

$$\begin{aligned}W^*(\pi^1_X,\pi^2_X) \leq &\tilde{\mathbb{E}}[c^*((\tilde{X}^1,\tilde{U},\tilde{B},\tilde{W}^1),(\tilde{X}^2,\tilde{U},\tilde{B},\tilde{W}^2))]\\ =&\tilde{\mathbb{E}}[\mathbb{1}_{\tilde{X}^1\neq\tilde{X}^2} + d(\tilde{W}^1,\tilde{W}^2)\wedge 1]\\ \leq&2.\end{aligned}$$

On the other hand, for any coupling of π^1 and π^2 such that $\mathbb{P}[(U^1,B^1)\neq(U^2,B^2)] > 0$, the quantity $\mathbb{E}[c^*((X^1,U^1,B^1,W^1),(X^2,U^2,B^2,W^2))] = \infty$. Therefore, the infimum (that is attained by some optimal coupling) in W^* may be taken over all couplings ensuring $\mathbb{P}[(U^1,B^1)\neq(U^2,B^2)] = 0$. By completing the probability space, it can be assumed that for the optimal coupling, $(U^1,B^1) = (U^2,B^2)$ surely and the superscripts will consequently be dropped.

To show that $W^*(\pi^1_X,\pi^2_X) = 0$, it will first be shown that $W^* = 0$ for solutions restricted to a short time interval. Define $p^i_{X,T}$ as the image of p^i_X through the map $\mathcal{C}\ni x\mapsto x_{\cdot\wedge T}\in\mathcal{C}$. Then, defining

$$\pi^i_{X,T} := p^i_{X,T}(dx, G^i(u,b), b, w)\lambda(du)\mathcal{W}_{d_B}(db)\mathcal{W}_{d_B}(dw),$$

see that for E^* equipped with $\pi^i_{X,T}$, and again defining $\mu^i := G^i(U,B)$, the elements X,μ,B,W have distribution $\mathscr{L}^i(X^i_{\cdot\wedge T},\mu^i,B^i,W^i)$. It will be shown that for some small T, $W^*_T := W^*(\pi^1_{X,T},\pi^2_{X,T}) = 0$ by representing the two measures via Girsanov transformations from the optimal coupling for W^*_T. Then, by repeating the argument, $W^*(\pi^1_X,\pi^2_X) = 0$ will be established by induction on intervals $[0,kT]$. The optimal coupling for W^*_T, denoted $\mathbb{P}$ henceforth, satisfies $X^i = X^i_{\cdot\wedge T}$ and for all $t\leq T$,

$$\begin{aligned}\mathbb{E}[d_{TV}(\mu^1_t,\mu^2_t)] \leq \mathbb{E}[\mathbb{E}[\mathbb{1}_{X^1_{\cdot\wedge t}\neq X^2_{\cdot\wedge t}}|\mathcal{F}^{B,U}]] = \mathbb{E}[\mathbb{1}_{X^1_{\cdot\wedge t}\neq X^2_{\cdot\wedge t}}] \leq &\mathbb{E}[\mathbb{1}_{X^1_{\cdot\wedge T}\neq X^2_{\cdot\wedge T}}]\\ =&W^*(\pi^1_{X,T},\pi^2_{X,T}).\end{aligned} \tag{4.3.8}$$

The following argument shows that for small T, $W^*_T = 0$:

4. By the Kantorovich Duality (see Theorem 5.10 in [100]), the primal and dual Kantorovich problems for c^* satisfy,

$$\begin{aligned}&W^*(\pi^1_{X,T},\pi^2_{X,T})\\ &= \sup_{h\ c^*\text{-convex}}\left(\int h(x,u,b,w)(\pi^1_{X,T}-\pi^2_{X,T})(dx,du,db,dw)\right) \qquad (4.3.9)\\ &= \sup_{h\ c^*\text{-convex}} \mathbb{E}[h(X^1,U,B,W^1) - h(X^2,U,B,W^2)]\end{aligned}$$

The second equality holds since $\mathbb{P}$ is a coupling of $\pi^1_{X,T}$ and $\pi^2_{X,T}$. The definition of c^* convexity, can be found in [100] p.54, but for the purposes here it will suffice to consider the equivalence that, since c^* satisfies the triangle inequality, h is c^*-convex iff

$$h(x^1, u^1, b^1, w^1) - h(x^2, u^2, b^2, w^2) \leq c^*((x^1, u^1, b^1, w^1), (x^2, u^2, b^2, w^2)). \tag{4.3.10}$$

It will be necessary to consider an alternative, but equivalent supremum in the right hand side of Equation (4.3.9), where one is able to assume that all functions h in the supremum are non-negative and bounded. This will be arrived at by the subsequent argument.

By the characterisation of c^*-convex functions, (4.3.10), for arbitrary but fixed $x' \in \mathcal{C}$ and $w' \in C(I; \mathbb{R}^{d_W})$, mapping every c^*-convex function h to a new c^*-convex function h' such that

$$h'(x, u, b, w) := h(x, u, b, w) - h(x', u, b, w') \leq c^*((x, u, b, w), (x', u, b, w')),$$

one can see that since c^* is symmetric, $|h'| \leq 2$. Finally, setting $h'' := h' + 2$ (again h'' is c^*-convex), see that for every c^*-convex h,

$$\begin{aligned}&\mathbb{E}[h(X^1, U, B, W^1) - h(X^2, U, B, W^2)] \\ &= \mathbb{E}[h''(X^1, U, B, W^1) - h''(X^2, U, B, W^2)]\end{aligned}$$

and h'' is $[0, 4]$ valued. Therefore, by sending every h to its corresponding h'',

$$W^*(\pi^1_{X,T}, \pi^2_{X,T}) = \sup_{h:E^* \to [0,4],\ c^*\text{-convex}} \mathbb{E}[h(X^1, U, B, W^1) - h(X^2, U, B, W^2)]. \tag{4.3.11}$$

5. Now, on the optimal probability space $(\Omega, \mathcal{F}, \mathbb{P})$, enlarged to include another Brownian motion W^0 (this is not necessary, since one could use W^1 or W^2 in place of W^0, but arguably this eases notation), there is a strong solution X^0 to the driftless equation (4.3.1) by Assumption 4.3.1. Indeed, there is a process X^0 such that

$$dX^0_t = \sigma(t, X^0)dW^0_t + \rho(t, X^0)dB_t.$$

In order to estimate the right hand side of (4.3.11), it is critical to represent the distributions of $X^i_{\cdot\wedge T}$ by the distributions of $X^0_{\cdot\wedge T}$ under suitable Girsanov transformations. For each $i = 1, 2$, define measures $\mathbb{Q}^i \sim \mathbb{P}$ by

$$\frac{d\mathbb{Q}^i}{d\mathbb{P}} := \mathcal{E}\left(\int_0^{\cdot\wedge T} \sigma^{-1}(s, X^0)b(s, X^0, \mu^i)dW^0_s\right)_\infty. \tag{4.3.12}$$

$\mathcal{E}(M)_t$ denotes the Doléans-Dade exponential of M at time t, $\mathcal{E}(M)_t := \exp\{M_t - \frac{1}{2}[M]_t\}$. These changes of probability measure are well defined due to the assumption of boundedness of $\sigma^{-1}b$. By Girsanov's Theorem, $W^{0,i} := W^0 - \int_0^{\cdot\wedge T} \sigma^{-1}(s, X^0)b(s, X^0, \mu^i)ds$ is a $\mathbb{Q}^i$ Brownian motion on I, and on $[0, T]$ and for each $i = 1, 2$,

$$dX^0_t = b(t, X^0, \mu^i)dt + \sigma(t, X^0)dW^{0,i}_t + \rho(t, X^0)dB_t.$$

It is now claimed that, $\mathscr{L}^i(X^0_{\cdot\wedge T}, U, B, W^{0,i}) = \mathscr{L}(X^i_{\cdot\wedge T}, U, B, W^i)$, where $\mathscr{L}^i$ denotes the law on $\mathbb{Q}^i$ (and continues to do so for the remainder of the proof). This follows from the uniqueness in joint law on $[0,T]$ for solutions for SDEs with random coefficients of the form:

$$dY_t = b(t, Y, \mu)dt + \sigma(t, Y)dW_t + \rho(t, Y)dB_t, \tag{4.3.13}$$

where the joint distribution of (μ, B, W) is determined. This uniqueness is given by Lemma 4.3.5, which is stated and proved at the end of the current proof.

6. Recalling the equation (4.3.11), and the two equivalent probability spaces $\mathbb{Q}^1$ and $\mathbb{Q}^2$,

$$\begin{aligned}
&W^*(\pi^1_{X,T}, \pi^2_{X,T})\\
=& \sup_{h:E^*\to[0,4],\text{ c-convex}} \mathbb{E}[h(X^1, U, B, W^1) - h(X^2, U, B, W^2)]\\
=& \sup_{h:E^*\to[0,4],\text{ c-convex}} \mathbb{E}^1[h(X^0_{\cdot\wedge T}, U, B, W^{0,1})] - \mathbb{E}^2[h(X^0_{\cdot\wedge T}, U, B, W^{0,2})]\\
=& \sup_{h:E^*\to[0,4],\text{ c-convex}} \mathbb{E}\left[\frac{d\mathbb{Q}^1}{d\mathbb{P}} h(X^0_{\cdot\wedge T}, U, B, W^{0,1}) - \frac{d\mathbb{Q}^2}{d\mathbb{P}} h(X^0_{\cdot\wedge T}, U, B, W^{0,2})\right]\\
=& \sup_{h:E^*\to[0,4],\text{ c-convex}} \left\{\mathbb{E}\left[\frac{d\mathbb{Q}^1}{d\mathbb{P}}\left(h(X^0_{\cdot\wedge T}, U, B, W^{0,1}) - h(X^0_{\cdot\wedge T}, U, B, W^{0,2})\right)\right]\right.\\
&\left. + \mathbb{E}\left[\left(\frac{d\mathbb{Q}^1}{d\mathbb{P}} - \frac{d\mathbb{Q}^2}{d\mathbb{P}}\right) h(X^0_{\cdot\wedge T}, U, B, W^{0,2})\right]\right\}
\end{aligned} \tag{4.3.14}$$

The right hand side of (4.3.14) will be estimated as follows:

$$\begin{aligned}
& \sup_{h:E^*\to[0,4],\text{ c-convex}} \left\{\mathbb{E}\left[\frac{d\mathbb{Q}^1}{d\mathbb{P}}\left(h(X^0_{\cdot\wedge T}, U, B, W^{0,1}) - h(X^0_{\cdot\wedge T}, U, B, W^{0,2})\right)\right]\right.\\
&\left. + \mathbb{E}\left[\left(\frac{d\mathbb{Q}^1}{d\mathbb{P}} - \frac{d\mathbb{Q}^2}{d\mathbb{P}}\right) h(X^0_{\cdot\wedge T}, U, B, W^{0,2})\right]\right\}\\
\leq & \sup_{h:E^*\to[0,4],\text{ c-convex}} \mathbb{E}^1[(h(X^0_{\cdot\wedge T}, U, B, W^{0,1}) - h(X^0_{\cdot\wedge T}, U, B, W^{0,2}))]\\
+ & \sup_{h:E^*\to[0,4],\text{ measurable}} \mathbb{E}^1\left[\left(1 - \frac{d\mathbb{Q}^2}{d\mathbb{P}^1}\right) h(X^0_{\cdot\wedge T}, U, B, W^{0,2})\right]
\end{aligned}$$

$$\begin{aligned}
&\leq \mathbb{E}^1[d(W^{0,1}, W^{0,2}) \wedge 1] + 4\mathbb{E}^1\left[\left(1 - \frac{d\mathbb{Q}^2}{d\mathbb{Q}^1}\right)\mathbb{1}_{\frac{d\mathbb{Q}^2}{d\mathbb{Q}^1}<1}\right]\\
&\leq \mathbb{E}^1[d(W^{0,1}, W^{0,2})] + 4\mathbb{E}^1\left[\left|1 - \frac{d\mathbb{Q}^2}{d\mathbb{Q}^1}\right|\mathbb{1}_{\frac{d\mathbb{Q}^2}{d\mathbb{Q}^1}<1}\right]
\end{aligned} \tag{4.3.15}$$

Recalling the definitions of W^i and the form of $\frac{d\mathbb{Q}^1}{d\mathbb{P}}$ and $\frac{d\mathbb{Q}^2}{d\mathbb{P}}$ from (4.3.12), $\frac{d\mathbb{Q}^2}{d\mathbb{Q}^1}$

can be rewritten as follows:

$$\begin{aligned}
&\frac{d\mathbb{Q}^2}{d\mathbb{Q}^1}\\
=&\exp\bigg\{\int_0^T \sigma^{-1}(s,X^0)b(s,X^0,\mu^2)dW_s^0-\int_0^T \sigma^{-1}(s,X^0)b(s,X^0,\mu^1)dW_s^0\\
&+\frac{1}{2}\int_0^T |\sigma^{-1}(s,X^0)b(s,X^0,\mu^1)|^2ds-\frac{1}{2}\int_0^T |\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)|^2ds\bigg\}\\
=&\exp\bigg\{\int_0^T \sigma^{-1}(s,X^0)b(s,X^0,\mu^2)dW_s^{0,1}-\int_0^T \sigma^{-1}(s,X^0)b(s,X^0,\mu^1)dW_s^{0,1}\\
&-\frac{1}{2}\int_0^T |\sigma^{-1}(s,X^0)b(s,X^0,\mu^1)|^2ds-\frac{1}{2}\int_0^T |\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)|^2ds\\
&+\int_0^T \sigma^{-1}(s,X^0)b(s,X^0,\mu^1)\cdot\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)ds\bigg\}\\
=&\exp\bigg\{-\int_0^T \sigma^{-1}(s,X^0)b(s,X^0,\mu^1)-\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)dW_s^{0,1}\\
&-\frac{1}{2}\int_0^T |\sigma^{-1}(s,X^0)b(s,X^0,\mu^1)-\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)|^2ds\bigg\}.
\end{aligned}\tag{4.3.16}$$

Now, on the event $\frac{d\mathbb{Q}^2}{d\mathbb{Q}^1}<1$,

$$\begin{aligned}
&\exp\bigg\{-\int_0^T \sigma^{-1}(s,X^0)b(s,X^0,\mu^1)-\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)dW_s^{0,1}\\
&-\frac{1}{2}\int_0^T |\sigma^{-1}(s,X^0)b(s,X^0,\mu^1)-\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)|^2ds\bigg\}<1.
\end{aligned}$$

Since for all $x\le 0$ (i.e. $e^x<1$), $|1-e^x|\le|x|$,

$$\begin{aligned}
&\mathbb{E}^1\bigg[\bigg|1-\frac{d\mathbb{Q}^2}{d\mathbb{Q}^1}\bigg|\mathbb{1}_{\frac{d\mathbb{Q}^2}{d\mathbb{Q}^1}<1}\bigg]\\
\le&\mathbb{E}^1\bigg[\bigg|-\int_0^T \sigma^{-1}(s,X^0)b(s,X^0,\mu^1)-\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)dW_s^{0,1}\\
&-\frac{1}{2}\int_0^T |\sigma^{-1}(s,X^0)b(s,X^0,\mu^1)-\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)|^2ds\bigg|\mathbb{1}_{\frac{d\mathbb{Q}^2}{d\mathbb{Q}^1}<1}\bigg]\\
\le&\mathbb{E}^1\bigg[\bigg|\int_0^T \sigma^{-1}(s,X^0)b(s,X^0,\mu^1)-\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)dW_s^{0,1}\bigg|\\
&+\frac{1}{2}\int_0^T |\sigma^{-1}(s,X^0)b(s,X^0,\mu^1)-\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)|^2ds\bigg]\\
\le&\mathbb{E}^1\bigg[\sup_{t\le T}\bigg|\int_0^t \sigma^{-1}(s,X^0)b(s,X^0,\mu^1)-\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)dW_s^{0,1}\bigg|\bigg]\\
&+\frac{1}{2}\mathbb{E}^1\bigg[\int_0^T |\sigma^{-1}(s,X^0)b(s,X^0,\mu^1)-\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)|^2ds\bigg].
\end{aligned}$$

Applying the Burkhölder-Davis-Gundy inequality (the corresponding constant denoted c_{BDG}),

$$\begin{aligned}
&\mathbb{E}^1\left[\left|1-\frac{d\mathbb{Q}^2}{d\mathbb{Q}^1}\right|\mathbb{1}_{\frac{d\mathbb{Q}^2}{d\mathbb{Q}^1}<1}\right]\\
\leq &c_{\mathsf{BDG}}\mathbb{E}^1\left[\left(\int_0^T|\sigma^{-1}(s,X^0)b(s,X^0,\mu^1)-\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)|^2ds\right)^{\frac{1}{2}}\right]\\
&+\frac{1}{2}\mathbb{E}^1\left[\int_0^T|\sigma^{-1}(s,X^0)b(s,X^0,\mu^1)-\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)|^2ds\right]
\end{aligned}$$

Now, using the assumption of total variation Lipschitz continuity of $\sigma^{-1}b$ in the measure component,

$$\begin{aligned}
&c_{\mathsf{BDG}}\mathbb{E}^1\left[\left(\int_0^T|\sigma^{-1}(s,X^0)b(s,X^0,\mu^1)-\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)|^2ds\right)^{\frac{1}{2}}\right]\\
&+\frac{1}{2}\mathbb{E}^1\left[\int_0^T|\sigma^{-1}(s,X^0)b(s,X^0,\mu^1)-\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)|^2ds\right]\\
\leq &c_{\mathsf{BDG}}c_{\mathsf{TV}}\mathbb{E}^1\left[\left(\int_0^T d_{\mathsf{TV}}(\mu_s^1,\mu_s^2)^2ds\right)^{\frac{1}{2}}\right]+\frac{1}{2}c_{\mathsf{TV}}^2\mathbb{E}^1\left[\int_0^T d_{\mathsf{TV}}(\mu_s^1,\mu_s^2)^2ds\right].
\end{aligned} \tag{4.3.17}$$

And since for all $s\leq T$, $d_{\mathsf{TV}}(\mu_s^1,\mu_s^2)\leq d_{\mathsf{TV}}(\mu_T^1,\mu_T^2)$,

$$\begin{aligned}
&\mathbb{E}^1\left[\left|1-\frac{d\mathbb{P}^2}{d\mathbb{P}^1}\right|\mathbb{1}_{\frac{d\mathbb{P}^2}{d\mathbb{P}^1}<1}\right]\\
\leq &c_{\mathsf{BDG}}c_{\mathsf{TV}}T^{\frac{1}{2}}\mathbb{E}^1[d_{\mathsf{TV}}(\mu_T^1,\mu_T^2)]+\frac{1}{2}c_{\mathsf{TV}}^2T\mathbb{E}^1[d_{\mathsf{TV}}(\mu_T^1,\mu_T^2)^2]\\
= &c_{\mathsf{BDG}}c_{\mathsf{TV}}T^{\frac{1}{2}}\mathbb{E}[d_{\mathsf{TV}}(\mu_T^1,\mu_T^2)]+\frac{1}{2}c_{\mathsf{TV}}^2T\mathbb{E}[d_{\mathsf{TV}}(\mu_T^1,\mu_T^2)^2]\\
\leq &c_{\mathsf{BDG}}c_{\mathsf{TV}}T^{\frac{1}{2}}\mathbb{E}[\mathbb{E}[\mathbb{1}_{X^1_{\cdot\wedge T}\neq X^2_{\cdot\wedge T}}|U,B]]+\frac{1}{2}c_{\mathsf{TV}}^2T\mathbb{E}[\mathbb{E}[\mathbb{1}_{X^1_{\cdot\wedge T}\neq X^2_{\cdot\wedge T}}|U,B]^2]\\
\leq &(c_{\mathsf{BDG}}c_{\mathsf{TV}}T^{\frac{1}{2}}+\frac{1}{2}c_{\mathsf{TV}}^2T)\mathbb{E}[\mathbb{E}[\mathbb{1}_{X^1_{\cdot\wedge T}\neq X^2_{\cdot\wedge T}}|U,B]]\\
= &(c_{\mathsf{BDG}}c_{\mathsf{TV}}T^{\frac{1}{2}}+\frac{1}{2}c_{\mathsf{TV}}^2T)\mathbb{P}[X^1_{\cdot\wedge T}\neq X^2_{\cdot\wedge T}]\\
= &(c_{\mathsf{BDG}}c_{\mathsf{TV}}T^{\frac{1}{2}}+\frac{1}{2}c_{\mathsf{TV}}^2T)W^*(\pi^1_{X,T},\pi^2_{X,T}).
\end{aligned}$$

Similarly, for $\mathbb{E}^1[d(W^{0,1},W^{0,2})]$, one estimates

$$\begin{aligned}
&\mathbb{E}^1[d(W^{0,1},W^{0,2})]\\
\leq &\mathbb{E}^1\left[\sup_{t\leq T}\left|\int_0^t\sigma^{-1}(s,X^0)b(s,X^0,\mu^1)-\sigma^{-1}(s,X^0)b(s,X^0,\mu^2)ds\right|\right]\\
\leq &c_{\mathsf{TV}}T\mathbb{E}^1[d_{\mathsf{TV}}(\mu_T^1,\mu_T^2)]\\
\leq &c_{\mathsf{TV}}TW^*(\pi^1_{X,T},\pi^2_{X,T}).
\end{aligned}$$

Putting the above two estimates together with (4.3.15),

$$W^*(\pi^1_{X,T},\pi^2_{X,T}) \leq \left(c_{\mathrm{TV}}T + 4\left(c_{\mathrm{BDG}}c_{\mathrm{TV}}T^{\frac{1}{2}} + \frac{1}{2}c^2_{\mathrm{TV}}T\right)\right)W^*(\pi^1_{X,T},\pi^2_{X,T}).$$

Hence, choosing T small enough such that $c_{\mathrm{TV}}T + 4(c_{\mathrm{BDG}}c_{\mathrm{TV}}T^{\frac{1}{2}} + \frac{1}{2}c^2_{\mathrm{TV}}T) = \alpha < 1$, one has

$$W^*(\pi^1_{X,T},\pi^2_{X,T}) \leq \alpha W^*(\pi^1_{X,T},\pi^2_{X,T}).$$

This implies that $W^*(\pi^1_{X,T},\pi^2_{X,T}) = 0$. Importantly, this further implies that almost surely, $G^1(U,B)_{\cdot\wedge T} = G^2(U,B)_{\cdot\wedge T}$. Indeed, since $G^i(U,B)_t = \mu^i_t = \mathscr{L}(X^i_{\cdot\wedge t}|U,B)$, for any $t\leq T$, and any $A\in\mathcal{B}(\mathcal{C})$,

$$\begin{aligned}\mathbb{E}[\mu^1_t(A)f(U,B)] = \mathbb{E}[\mathbb{1}_A(X^1_{\cdot\wedge t})f(U,B)] =& \mathbb{E}[\mathbb{1}_A(X^2_{\cdot\wedge t})f(U,B)]\\ =& \mathbb{E}[\mu^2_{\cdot\wedge t}(A)f(U,B)].\end{aligned}$$

This means that the distribution of $(G^1(U,B)_{\cdot\wedge T}, G^2(U,B)_{\cdot\wedge T})$ is concentrated on the diagonal (and will be on any probability space supporting (U,B) with the same distribution).

7. The result of the proof will follow by an inductive argument. Assume that for some $k\in\mathbb{N}$ $W^*(\pi^1_{X,kT},\pi^2_{X,kT}) = 0$, then repeating the above argument for $\pi^1_{X,(k+1)T}$ and $\pi^2_{X,(k+1)T}$, then, since $\mu^1 = \mu^2$ almost surely on $[0,kT]$,

$$\begin{aligned}&W^*(\pi^1_{X,(k+1)T},\pi^2_{X,(k+1)T})\\ \leq& 4c_{\mathrm{BDG}}c_{\mathrm{TV}}\mathbb{E}^1\left[\left(\int_0^{(k+1)T} d_{\mathrm{TV}}(\mu^1_s,\mu^2_s)^2ds\right)^{\frac{1}{2}}\right]\\ &+ 4\frac{1}{2}c^2_{\mathrm{TV}}\mathbb{E}^1\left[\int_0^{(k+1)T} d_{\mathrm{TV}}(\mu^1_s,\mu^2_s)^2ds\right]\\ &+ c_{\mathrm{TV}}\mathbb{E}^1\left[\int_0^{(k+1)T} d_{\mathrm{TV}}(\mu^1_s,\mu^2_s)ds\right]\\ =& 4c_{\mathrm{BDG}}c_{\mathrm{TV}}\mathbb{E}^1\left[\left(\int_{kT}^{(k+1)T} d_{\mathrm{TV}}(\mu^1_s,\mu^2_s)^2ds\right)^{\frac{1}{2}}\right]\\ &+ 4\frac{1}{2}c^2_{\mathrm{TV}}\mathbb{E}^1\left[\int_{kT}^{(k+1)T} d_{\mathrm{TV}}(\mu^1_s,\mu^2_s)^2ds\right]\\ &+ c_{\mathrm{TV}}\mathbb{E}^1\left[\int_{kT}^{(k+1)T} d_{\mathrm{TV}}(\mu^1_s,\mu^2_s)ds\right]\\ \leq& \left(c_{\mathrm{TV}}T + 4\left(c_{\mathrm{BDG}}c_{\mathrm{TV}}T^{\frac{1}{2}} + \frac{1}{2}c^2_{\mathrm{TV}}T\right)\right)W^*(\pi^1_{X,(k+1)T},\pi^2_{X,(k+1)T}).\end{aligned}$$

Therefore $W^*(\pi^1_{X,(k+1)T},\pi^2_{X,(k+1)T}) = 0$. By induction, the proof is complete.

□

Chapter 5

Extensions and Future Work

There are a few directions in which the work contained in this thesis could be extended:

1. The Lyapunov criteria from Chapter 2 are expected to be generalisable to the common noise setting. After applying the corresponding, more involved Itô formula for this setting (see Section 4.3 of Volume II of [26]) to a candidate Lyapunov function, by use of the tower property, one sees that a condition such as Assumption 2.2.2 remains appropriate in this setting.

2. It seems that one might be able to prove that the conditional propagation of chaos occurs under the conditions of Theorem 4.3.3, as such a result is established under similar conditions in the absence of a common noise in [70]. However, there are significant obstacles in the common noise setting.

3. In [84], Mishura and Veretennikov allow for integration kernels that are locally bounded, satisfying a linear growth condition in the argument that is not integrated over. One possible extension of the results in Chapter 4 would be to allow for linear growth in the spatial variable.

Appendix A

Measure Derivatives of Lions and Associated Itô Formula

For the construction of the measure derivative in the sense of Lions we follow the approach from [22, Section 6]. There are three main differences: The first difference is that we define the measure derivative in a domain. More precisely we will define the measure derivative for any measure as long as it has support on $D_k \subset D$ for some $k \in \mathbb{N}$ (recall that $\overline{D}_k \subset D_{k+1}$ and $\bigcup_k D_k = D$ and every D_k is bounded and open), in practice for the processes x^k, this may be D_{k+1}.

This is precisely what is needed for the analysis in Chapter 2. The second difference is that we are explicit in making it clear why the measure derivative is independent of the probability space used to realise the measure as well as the random variable used. The third difference is in proving the "Structure of the gradient", see [22, Theorem 6.5]. Thanks to an observation by Sandy Davie (University of Edinburgh), we can show as part iii) of Proposition A.1.2 that the measure derivative has the right structure even if it only exists at the point μ instead of for every square integrable measure, as is required in [22]. The method of Sandy Davie also conveniently results in a much shorter proof. We assume the same regularity as in [27], but less regularity is assumed than in [19] following the observations of [77]

A.1 Lions' Measure Derivative on D_k

Consider $u : \mathcal{P}_2(D) \to \mathbb{R}$. Here $\mathcal{P}_2(D)$ is a space of probability measures on D that have second moments i.e. $\int_D x^2 \mu(dx) < \infty$ for $\mu \in \mathcal{P}_2(D)$. We want to define the derivative at points $\mu \in \mathcal{P}_2(D)$ such that $\operatorname{supp}(\mu) \subseteq D_k$. We shall write $\mu \in \mathcal{P}(D_k)$ if μ is a probability measure on D with support in D_k.

Definition A.1.1 (L-differentiability at $\mu \in \mathcal{P}(D_k)$). We say that u is L-differentiable at $\mu \in \mathcal{P}(D_k)$ if there is an atomless Polish probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an $X \in L^2(\Omega)$ such that $\mu = \mathscr{L}(X)$ and the function $U : L^2(\Omega) \to \mathbb{R}$ given by $U(Y) := u(\mathscr{L}(Y))$ is Fréchet differentiable at X. We will call U the *lift* of u.

We recall that saying $U : L^2(\Omega; D) \to \mathbb{R}$ is Fréchet differentiable at X with $\operatorname{supp}(X) \subseteq D_k$ means that there exists a bounded linear operator $A : L^2(\Omega) \to \mathbb{R}$

such that for

$$\lim_{\substack{|Y|_2\to 0\\ \mathrm{supp}(X+Y)\subseteq D}} \left| \frac{U(X+Y)-U(X)}{|Y|_2} - \frac{AY}{|Y|_2} \right| = 0\,. \tag{A.1.1}$$

Note that Since $L^2(\Omega)$ is a Hilbert space with the inner product $(X,Y) := \mathbb{E}[XY]$ we can identify $L^2(\Omega)$ with its dual $L^2(\Omega)^*$ via this inner product. Then the bounded linear operator A defines an element $DU(X) \in L^2(\Omega)$ through

$$(DU(X), Y) := AY \quad \forall Y \in L^2(\Omega).$$

Proposition A.1.2. *Let u be L-differentiable at $\mu \in \mathcal{P}(D_k)$, with some atomless $(\Omega, \mathcal{F}, \mathbb{P})$, lift U and $X \in L^2(\Omega)$ such that $\mu = \mathscr{L}(X)$. Let $(\bar\Omega, \bar{\mathcal{F}}, \bar{\mathbb{P}})$ be an arbitrary atomless, Polish probability space which supports $\bar X \in L^2(\bar\Omega)$ and on which we have the lift $\bar U(Y) := u(\mathscr{L}(Y))$. Then*

i) The lift $\bar U$ is Fréchet differentiable at $\bar X$ with derivative $D\bar U(\bar X) \in L^2(\bar\Omega)$.

ii) The joint law of $(X, DU(X))$ equals that of $(\bar X, D\bar U(\bar X))$.

iii) There is $\xi : D_k \to D_k$ measurable such that $\int_{D_k} \xi^2(x)\mu(dx) < \infty$ and almost surely,

$$\xi(X) = DU(X), \quad \xi(\bar X) = D\bar U(\bar X)\,.$$

Definition A.1.3 (L-derivative of u at μ)**.** If u is L-differentiable at μ then we write $\partial_\mu u(\mu) := \xi$, where ξ is given by Proposition A.1.2. Moreover we have $\partial_\mu u : \mathcal{P}_2(D_k) \times D_k \to D_k$ given by

$$\partial_\mu u(\mu, y) := [\partial_\mu u(\mu)](y)\,.$$

To prove Proposition A.1.2 we will need the following result:

Lemma A.1.4. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ and $(\bar\Omega, \bar{\mathcal{F}}, \bar{\mathbb{P}})$ be two atomless, Polish probability spaces supporting D_k-valued random variables X and $\bar X$ such that $\mathscr{L}(X) = \mathscr{L}(\bar X)$. Then for any $\epsilon > 0$ there exists $\tau : \Omega \to \bar\Omega$ which is bijective, such that both τ and τ^{-1} are measurable and measure preserving and moreover*

$$|X - \bar X \circ \tau|_\infty < \epsilon \quad \textit{and} \quad |X \circ \tau^{-1} - \bar X|_\infty < \epsilon\,.$$

Proof. Let $(A_n)_n$ be a measurable partition of D_k such that $\mathrm{diam}(A_n) < \epsilon$. Let

$$B_n := \{X \in A_n\}, \quad \bar B_n := \{\bar X \in A_n\}\,.$$

These form measurable partitions of Ω and $\bar\Omega$ respectively and moreover $\mathbb{P}(B_n) = \bar{\mathbb{P}}(\bar B_n)$. As the probability spaces are atomless, there exist $\tau_n : B_n \to \bar B_n$ bijective, such that τ_n and τ_n^{-1} are measurable and measure preserving. See [43, Sec. 41, Theorem C] for details. Let

$$\tau(\omega) := \tau_n(\omega) \text{ if } \omega \in B_n, \quad \tau^{-1}(\bar\omega) := \tau_n^{-1}(\bar\omega) \text{ if } \bar\omega \in \bar B_n\,.$$

We can see that these are measurable, measure preserving bijections. Now consider $\omega \in B_n$. Then $\tau(\omega) = \tau_n(\omega) \in \bar B_n$. But then $X(\omega) \in A_n$ and $\bar X(\tau(\omega)) \in A_n$ too. Hence

$$|X(\omega) - \bar X(\tau(\omega))| < \epsilon \quad \forall \omega \in \Omega\,.$$

The estimate for the inverse is proved analogously. □

We use the notation $L^2 := L^2(\Omega)$ and $\bar{L}^2 := L^2(\bar{\Omega})$.

Proof of Proposition A.1.2, part i).
For any $h > 0$ we have τ_h, τ_h^{-1} given by Lemma A.1.4 measure preserving and such that $|X - \bar{X} \circ \tau_h|_\infty < h$. This means that $|X - \bar{X} \circ \tau_h|_2 < h$ and we have the analogous estimate with τ_h^{-1}. Our first aim is to show that $(DU(X) \circ \tau_h^{-1})_{h>0}$ is a Cauchy sequence in $\bar{L}^2$.

Fix $\epsilon > 0$. Then $\exists \delta > 0$ such that we have

$$|U(X+Y)-U(X)-(DU(X),Y)| < \frac{\epsilon}{2}|Y|_2 \quad \text{for all } |Y|_2 < \delta \text{ and } \operatorname{supp}(X+Y) \subseteq D\,,$$

since U is Fréchet differentiable at X. Fix $h, h' < \delta/2$ and consider $|\bar{Y}|_2 < \delta/2$ and $\operatorname{supp}(\bar{X} + \bar{Y}) \subseteq D$. Then, since the maps τ_h^{-1} are measure preserving, we have

$$(DU(X) \circ \tau_h^{-1}, \bar{Y}) = (DU(X), \bar{Y} \circ \tau_h)\,.$$

Note that the inner product on the left is in $\bar{L}^2$ but the one on the right is in L^2. This will not be distinguished in our notation. Let $Z_h := \bar{Y} \circ \tau_h - X + \bar{X} \circ \tau_h$. Then $|Z_h|_2 \leq |\bar{Y}|_2 + |\bar{X} \circ \tau_h - X|_2 < \delta$ and since $\operatorname{supp}(\bar{X} + \bar{Y}) \subseteq D$, we have $\operatorname{supp}(X + Z_h) \subseteq D$. Moreover

$$\begin{aligned}
(DU(X) \circ \tau_h^{-1} - DU(X) \circ \tau_{h'}^{-1}, \bar{Y}) &= (DU(X), Z_h) - (DU(X), Z_{h'}) \\
&\quad + (DU(X), \bar{X} \circ \tau_h - X) + (DU(X), X - \bar{X} \circ \tau_{h'}) \\
&= - U(X + Z_h) + U(X) + (DU(X), Z_h) + [U(X + Z_h) - U(X)] \\
&\quad + U(X + Z_{h'}) - U(X) - (DU(X), Z_{h'}) - [U(X + Z_{h'}) - U(X)] \\
&\quad + (DU(X), \bar{X} \circ \tau_h - X) + (DU(X), X - \bar{X} \circ \tau_{h'})\,.
\end{aligned}$$

But as τ_h is measure preserving and U and $\bar{U}$ only depend on the law, we have

$$U(X + Z_h) = U(\bar{Y} \circ \tau_h + \bar{X} \circ \tau_h) = \bar{U}(\bar{Y} + \bar{X}) = U(X + Z_{h'}).$$

Hence

$$\begin{aligned}
|(DU(X) \circ \tau_h^{-1} - DU(X) \circ \tau_{h'}^{-1}, \bar{Y})| &\leq \frac{\epsilon}{2}|Z_{h'}|_2 + \frac{\epsilon}{2}|Z_h|_2 + 2|DU(X)|_2 \max(h, h') \\
&\leq \epsilon|Y|_2 + \epsilon \max(h, h') + 2|DU(X)|_2 \max(h, h')\,.
\end{aligned}$$

This means that

$$\begin{aligned}
&|DU(X) \circ \tau_h^{-1} - DU(X) \circ \tau_{h'}^{-1}|_2 \\
&= \sup_{|\bar{Y}|_2 = \delta/2} \frac{|(DU(X) \circ \tau_h^{-1} - DU(X) \circ \tau_{h'}^{-1}, \bar{Y})|}{|\bar{Y}|_2} \leq \epsilon + (2\epsilon + 4|DU(X)|_2)\frac{\max(h, h')}{\delta}\,.
\end{aligned}$$

Since we can choose $h, h' < \frac{\delta}{2}$ and also $h, h' < \frac{\epsilon\delta}{4|DU(X)|_2}$ we have the required estimate and see that $(DU(X) \circ \tau_h^{-1})_{h>0}$ is a Cauchy sequence in $\bar{L}^2$. Thus, there is $\psi \in \bar{L}^2$ such that

$$DU(X) \circ \tau_h^{-1} \to \psi \quad \text{as} \quad h \searrow 0.$$

The next step is to show that $\bar{U}$ is Fréchet differentiable at $\bar{X}$ and $\psi = D\bar{U}(\bar{X})$. To that end we note that $\bar{U}(\bar{X}+\bar{Y}) = U(X+Z_h)$ and

$$(DU(X), \bar{Y}\circ\tau_h) = (DU(X), Z_h) + (DU(X), X - \bar{X}\circ\tau_h).$$

Hence

$$\begin{aligned}
&|\bar{U}(\bar{X}+\bar{Y}) - \bar{U}(\bar{X}) - (\psi,\bar{Y})| \\
&= |U(X+Z_h) - U(X) - (DU(X),\bar{Y}\circ\tau_h) + (DU(X),\bar{Y}\circ\tau_h) - (\psi,\bar{Y})| \\
&\leq \varepsilon|Z_h|_2 + |DU(X)|_2 h + |DU(X)\circ\tau_h^{-1} - \psi||\bar{Y}|_2 \leq 4\epsilon|\bar{Y}|_2,
\end{aligned}$$

for h sufficiently small. Thus $\bar{U}$ is differentiable at $\bar{X}$ and $\psi = D\bar{U}(\bar{X}) \in \bar{L}^2$. □

Proof of Proposition A.1.2, part ii). We first note that

$$\mathscr{L}(X\circ\tau_h^{-1}, DU(X)\circ\tau_h^{-1}) = \mathscr{L}(X, DU(X))$$

since the mapping τ_h^{-1} is measure preserving. Moreover

$$(X\circ\tau_h^{-1}, DU(X)\circ\tau_h^{-1}) \to (\bar{X}, D\bar{U}((\bar{X}))) \text{ in } L^2(\bar{\Omega};\mathbb{R}^{2d}) \text{ as } h\searrow 0.$$

Hence we get that $\mathscr{L}(X, DU(X)) = \mathscr{L}(\bar{X}, D\bar{U}(\bar{X}))$. □

Proof of Proposition A.1.2, part iii). Note that μ is not necessarily atomless. Let us first consider the case where it is. Then, equipping $(\tilde{\Omega},\tilde{\mathcal{F}}) := (D_k, \mathcal{B}(D_k))$ with $\tilde{\mathbb{P}} := \mu$, this probability space is atomless and the canonical random element $\tilde{X}(x) := x$ has law μ. Further, there is an $L^2(D_k)$ random element $D\tilde{U}(\tilde{X})$ that is the Fréchet derivative of the lift $\tilde{U}$ at $\tilde{X}$. Setting $\xi(x) = D\tilde{U}(\tilde{X})(x)$, then by the by uniqueness of the joint distribution, for any probability space $(\bar{\Omega},\bar{\mathcal{F}},\bar{\mathbb{P}})$ with $\bar{X}\in L^2(\bar{\Omega})$ s.t. $\mathscr{L}(\bar{X}) = \mu$, we have $\xi(\bar{X}) = D\bar{U}(\bar{X})$ almost surely.

To deal with the case where μ is not necessarily atomless, we take λ, the translation invariant measure on $(S^1,\mathcal{B}(S^1))$, with S^1 denoting the unit circle. Then, the probability space $(\tilde{\Omega},\tilde{\mathcal{F}},\tilde{\mathbb{P}}) := (D_k\times S^1, \mathcal{B}(D_k)\otimes\mathcal{B}(S^1), \mu\otimes\lambda)$ is atomless. Let $\tilde{L}^2$ denote the space of square integrable random variables on this probability space. The random variable $\tilde{X}(x,s) := x$ is in $\tilde{L}^2$ and has law μ. With the usual lift $\tilde{U}$ we know, from part i), that $D\tilde{U}(\tilde{X})$ exists in $\tilde{L}^2$.

For all $t\in S^1$, define the translation operator on $\tilde{L}^2$, by $R_t\tilde{Z}(x,s) = \tilde{Z}(x,s-t)$. Note that $R_t\tilde{X} = \tilde{X}$. Moreover, $\mathscr{L}(R_t\tilde{Z}) = \mathscr{L}(\tilde{Z})$ for all $\tilde{Z}\in\tilde{L}^2$ since, by translation invariance of λ,

$$\begin{aligned}
\int_{\tilde{\Omega}} \mathbb{1}_B(R_tZ(x,s))\lambda\otimes\mu(ds,dx) &= \int_{D_k}\int_{S^1}\mathbb{1}_B(Z(x,s-t))\lambda(ds)\mu(dx) \\
&= \int_{D_k}\int_{S^1}\mathbb{1}_B(Z(x,s))\lambda(ds)\mu(dx).
\end{aligned}$$

Since the lift $\tilde{U}(\tilde{Z})$ depends only on the distribution of the random element $\tilde{Z}$, $\tilde{U}(\tilde{Z}) = \tilde{U}(R_t\tilde{Z})$ and so $\tilde{U}(\tilde{X}+R_t\tilde{Z}) = \tilde{U}(\tilde{X}+\tilde{Z})$. Then by uniqueness of the Fréchet derivative, recalling equation (A.1.1), one can conclude that $\tilde{A} = \tilde{A}R_t$ on $\tilde{L}^2$, for all $t \in S^1$. Therefore, for all $t \in S^1$ and $\tilde{Y} \in \tilde{L}^2$, $\tilde{A}R_t(\tilde{Y}) = \tilde{A}(R_t\tilde{Y}) = \tilde{A}\tilde{Y}$ and so,

$$\begin{aligned}(D\tilde{U}(\tilde{X}), \tilde{Y}) &= \tilde{A}\tilde{Y} = \tilde{A}R_t\tilde{Y} = (D\tilde{U}(\tilde{X}), R_t\tilde{Y}) \\ &= \int_{D_k}\int_{S^1} D\tilde{U}(\tilde{X})(x,s)R_t\tilde{Y}(x,s)\lambda(ds)\mu(dx) \\ &= \int_{D_k}\int_{S^1} R_{-t}D\tilde{U}(\tilde{X})(x,s)\tilde{Y}(x,s)\lambda(ds)\mu(dx) = (R_{-t}D\tilde{U}(\tilde{X}), \tilde{Y}).\end{aligned}$$

Hence, $\tilde{\xi}(x,s) := D\tilde{U}(\tilde{X})(x,s)$ does not depend on s for x in the support of μ. Write $\xi(x) := \tilde{\xi}(x, s_0)$ for some $s_0 \in S^1$. Then,

$$1 = \mu \otimes \lambda \left(\xi(x) = D\tilde{U}(\tilde{X})(x,s)\right) = \tilde{\mathbb{P}}\left(\xi(\tilde{X}) = D\tilde{U}(\tilde{X})\right) = \bar{\mathbb{P}}\left(\xi(\bar{X}) = D\bar{U}(\bar{X})\right)$$

since $\mathscr{L}(\bar{X}, D\bar{U}(\bar{X})) = \mathscr{L}(\tilde{X}, D\tilde{U}(\tilde{X}))$ due to part ii) and $\xi(\bar{X}) = D\bar{U}(\bar{X})$ $\bar{\mathbb{P}}$-a.s. as required. □

A.2 Higher-Order Derivatives

We observe that if μ is fixed then $\partial_\mu u(\mu)$ is a function from $D_k \to D_k$. If, for $y \in D_k$, $\partial_y [\partial_\mu u(\mu)(y)_j]$ exists for each $j = 1, \ldots, d$ then $\partial_y\partial_\mu u : \mathcal{P}(D_k) \times D_k \to D_k \times D_k$ is the matrix

$$\partial_y\partial_\mu u(\mu, y) := \left(\partial_y [\partial_\mu u(\mu)(y)_j]\right)_{j=1,\ldots,d}.$$

If we fix $y \in D_k$ then $\partial_\mu u(\cdot)(y)$ is a function from $\mathcal{P}(D_k) \to D_k$. Fixing $j = 1, \ldots, d$, if $\partial_\mu u(\cdot)(y)_j : \mathcal{P}(D_k) \to \mathbb{R}$ is L-differentiable at some μ then its L-derivative is the function given by part iii) of Proposition A.1.2, namely $\partial_\mu\left(\partial_\mu u(\mu)(y)_j\right) : D_k \to D_k$. The second order derivative in measure thus constructed is $\partial^2_\mu : \mathcal{P}(D_k) \times D_k \times D_k \to D_k \times D_k$ given by

$$\partial^2_\mu(\mu, y, \bar{y}) := \left(\partial_\mu\left(\partial_\mu u(\mu, y)_j\right)(\bar{y})\right)_{j=1,\ldots,d}.$$

A.3 Itô Formula for Functions of Measures

Assume we have a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with filtration $(\mathcal{F}_t)_{t\geq 0}$ satisfying the usual conditions supporting an $(\mathcal{F}_t)_{t\geq 0}$-Brownian motion w and adapted processes b and σ satisfying appropriate integrability conditions. We consider the Itô process

$$dx_t = b_t\,dt + \sigma_t dw_t, \quad x_0 \in L^2(\mathcal{F}_0)$$

which satisfies $x_t \in D_k$ for all t a.s.

Definition A.3.1. We say that $u : \mathcal{P}_2(D) \to \mathbb{R}$ is in $\mathcal{C}^{(1,1)}(\mathcal{P}_2(D))$ if there is a continuous version of $y \mapsto \partial_\mu u(\mu)(y)$ such that the mapping $\partial_\mu u : \mathcal{P}_2(D) \times D \to D$ is jointly continuous at any (μ, y) s.t. $y \in \text{supp}(\mu)$ and such that $y \mapsto \partial_\mu u(\mu, y)$ is continuously differentiable and its derivative $\partial_y\partial_\mu u : \mathcal{P}_2(D) \times D \to D \times D$ is jointly continuous at any (μ, y) s.t. $y \in \text{supp}(\mu)$.

The notation $\mathcal{C}^{(1,1)}$ is chosen to emphasise that we can take one measure derivative which is again differentiable (in the usual sense) with respect to the new free variable that arises. Note that in [27] such functions are called partially $\mathcal{C}^2$.

Proposition A.3.2. *Assume that*

$$\mathbb{E}\left[\int_0^\infty |b_t|^2 + |\sigma_t|^4 \, dt\right] < \infty .$$

Let u be in $\mathcal{C}^{(1,1)}(\mathcal{P}_2(D))$ such that for any compact subset $\mathcal{K} \subset \mathcal{P}_2(D)$

$$\sup_{\mu \in \mathcal{K}} \int_D \left[|\partial_\mu u(\mu)(y)|^2 + |\partial_y \partial_\mu u(\mu)(y)|^2\right] \mu(dy) < \infty . \tag{A.3.1}$$

Then, for $\mu_t := \mathscr{L}(x_t)$,

$$u(\mu_t) = u(\mu_0) + \int_0^t \mathbb{E}\left[b_s \partial_\mu u(\mu_s)(x_s) + \frac{1}{2} tr\left[\sigma_s \sigma_s^T \partial_y \partial_\mu u(\mu_s)(x_s)\right]\right] ds .$$

Note that since we are assuming that the process x never leaves some D_k, we have $\mathrm{supp}(\mu_t) \subseteq D_k$ for all times t. The proof relies on replacing μ_t by an approximation arising as the empirical measure of N independent copies of the process x. For marginal empirical measures there is a direct link between measure derivatives and partial derivatives, see [27, Proposition 3.1]. One can then apply the classical Itô formula to the approximating system of independent copies of x and take the limit. This is done in [27, Theorem 3.5].

Proposition A.3.2 can be used to derive an Itô formula for a function which depends on (t, x, μ).

Definition A.3.3. By $\mathcal{C}^{1,2,(1,1)}([0,\infty) \times D \times \mathcal{P}_2(D))$ we denote the functions $v = v(t, x, \mu)$ such that $v(\cdot, \cdot, \mu) \in C^{1,2}([0,\infty) \times D)$ for each μ, and such that $v(t, x, \cdot)$ is in $\mathcal{C}^{(1,1)}(\mathcal{P}_2(D))$ for each (t, x). Moreover all the resulting (partial) derivatives must be jointly continuous in (t, x, μ) or (t, x, μ, y) as appropriate.

Finally, let $\mathcal{C}^{2,(1,1)}(D \times \mathcal{P}_2(D))$ denote the subspace of $\mathcal{C}^{1,2,(1,1)}([0,\infty) \times D \times \mathcal{P}_2(D))$ of functions v that are constant in t.

To conveniently express integrals with respect to the laws of the process taken only over the "new" variables arising in the measure derivative we introduce another probability space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}})$ a filtration $(\tilde{\mathcal{F}}_t)_{t \geq 0}$ and processes $\tilde{w}$, $\tilde{b}$, $\tilde{\sigma}$ and a random variable $\tilde{x}_0$ on this probability space such that they have the same laws as w, b, σ and x_0. We assume $\tilde{w}$ is a Wiener process. Then

$$d\tilde{x}_t = \tilde{b}_t \, dt + \tilde{\sigma}_t d\tilde{w}_t, \;\; \tilde{x}_0 \in L^2(\tilde{\mathcal{F}}_0)$$

is another Itô process which satisfies $\tilde{x}_t \in D_k$ for all t a.s. Moreover, if we now consider the probability space $(\Omega \times \tilde{\Omega}, \mathcal{F} \otimes \tilde{\mathcal{F}}, \mathbb{P} \otimes \tilde{\mathbb{P}})$ then we see that the processes with and without tilde are independent on this new space.

Proposition A.3.4 (Itô formula). *Assume that*

$$\mathbb{E}\left[\int_0^\infty |b_t|^2 + |\sigma_t|^4\, dt\right] < \infty\,.$$

Let $v \in \mathcal{C}^{1,2,(1,1)}([0,\infty)\times D\times \mathcal{P}_2(D))$ such that for any compact subset $\mathcal{K}\subset\mathcal{P}_2(D)$

$$\sup_{t,x,\mu\in[0,\infty)\times D\times\mathcal{K}} \int_D \left[|\partial_\mu v(t,x,\mu)(y)|^2 + |\partial_y\partial_\mu v(t,x,\mu)(y)|^2\right]\mu(dy) < \infty\,. \tag{A.3.2}$$

Then, for $\mu_t := \mathscr{L}(\tilde{x}_t)$,

$$\begin{aligned} v(t,x_t,\mu_t)-&v(0,x_0,\mu_0)\\ &= \int_0^t \left[\partial_t v(s,x_s,\mu_s) + b_s\partial_x v(s,x_s,\mu_s) + \frac{1}{2}tr\left[\sigma_s\sigma_s^T\partial_x^2 v(s,x_s,\mu_s)\right]\right] ds\\ &\quad + \int_0^t \sigma_s\partial_x v(s,x_s,\mu_s)\,dw_s\\ &\quad + \int_0^t \tilde{\mathbb{E}}\left[\tilde{b}_s\partial_\mu v(s,x_s,\mu_s)(\tilde{x}_s) + \frac{1}{2}tr\left[\tilde{\sigma}_s\tilde{\sigma}_s^T\partial_y\partial_\mu v(s,x_s,\mu_s)(\tilde{x}_s)\right]\right] ds\,. \end{aligned}$$

Here we follow the argument from [19] explaining how to go from an Itô formula for function of measures only, i.e. from Proposition A.3.2, to the general case. Note that it is possible to assume that $\tilde{w}$, $\tilde{b}$, $\tilde{\sigma}$ and $\tilde{x}_0$ have the same laws as w, b, σ as x_0 above, but in fact this is not necessary. In this thesis, this generality is needed in the proof of Lemma 2.2.11.

Outline of proof for Proposition A.3.4. Fix $(\bar{t},\bar{x})$ and apply Proposition A.3.2 to the function $u(\mu) := v(\bar{t},\bar{x},\mu)$ and the law $\mu_t := \mathscr{L}(\tilde{x}_t)$. Then

$$\begin{aligned} v(\bar{t},\bar{x},\mu_t)-&v(\bar{t},\bar{x},\mu_0)\\ &= \int_0^t \tilde{\mathbb{E}}\left[\tilde{b}_s\partial_\mu v(\bar{t},\bar{x},\mu_s)(\tilde{x}_s) + \frac{1}{2}\mathrm{tr}\left[\tilde{\sigma}_s\tilde{\sigma}_s^T\partial_y\partial_\mu v(\bar{t},\bar{x},\mu_s)(\tilde{x}_s)\right]\right] ds\\ &=: \int_0^t M(\bar{t},\bar{x},\mu_s)\,ds\,. \end{aligned}$$

We thus see that the map $t\mapsto v(\bar{t},\bar{x},\mu_t)$ is absolutely continuous for all $(\bar{t},\bar{x})$ and so for almost all t we have $\partial_t v(\bar{t},\bar{x},\mu_t) = M(\bar{t},\bar{x},\mu_t)$. Note that for completeness we would need to use the definition of $\mathcal{C}^{1,2,(1,1)}$ functions and a limiting argument to get the partial derivative for all t. See the proof of the corresponding Itô formula in [27]. We now consider $\bar{v}$ given by $\bar{v}(t,x) := v(t,x,\mu_t)$. Then $\partial_t\bar{v}(t,x) = (\partial_t v)(t,x,\mu_t) + M(t,x,\mu_t)$. Using the usual Itô formula we then have

$$\begin{aligned} \bar{v}(t,x_t) - \bar{v}(0,x_0) = &\int_0^t \left[\partial_t v(s,x_s,\mu_s) + M(s,x_s,\mu_s) + \frac{1}{2}\mathrm{tr}\left[\sigma_t\sigma_t^T\partial_x^2 v(s,x_s,\mu_s)\right]\right] ds\\ &+ \int_0^t b_s\partial_x v(s,x_s,\mu_s)\,dw_s\,. \end{aligned}$$

□

Appendix B

Appendix to Chapters 3 and 4

The following lemma is standard and numerous lemmas of this type are proved in the note [95].

Lemma B.0.1 (Doob-Dynkin Lemma). *Given measurable spaces* $(\Omega, \mathcal{F})$, $(\mathcal{X}, \mathcal{F}_{\mathcal{X}})$ *and* $(\mathcal{Y}, \mathcal{F}_{\mathcal{Y}})$, *with measurable functions* $X : \Omega \mapsto \mathcal{X}$ *and* $Y : \Omega \mapsto \mathcal{Y}$, *if the image* $X(\Omega)$ *of function* X *is contained in a standard Borel space, and* X *is measurable with respect to the initial* σ*-algebra of* Y *(the initial sigma algebra of* Y *is defined as* $\sigma(Y^{-1}(A) : A \in \mathcal{F}_{\mathcal{Y}})$*), then there exists a measurable* $\phi : \mathcal{Y} \mapsto \mathcal{X}$ *such that* $X = \phi(Y)$.

B.1 Immersion and Compatibility

The following theorem comes from [7] where further equivalent conditions and references can be found.

Theorem B.1.1 (Conditions equivalent to Immersion). *On a given probability space* $(\Omega, \mathcal{F}, \mathbb{P})$, *consider two filtrations* $\mathbb{F}, \mathbb{G}$ *such that* $\mathbb{F} \subset \mathbb{G}$. *Then* $\mathbb{F}$ *is immersed in* $\mathbb{G}$ *under* $\mathbb{P}$ *if and only if any of the following conditions holds:*

1. $\mathcal{G}_t$ *is conditionally independent of* $\mathcal{F}_\infty$ *given* $\mathcal{F}_t$, *for any* t.

2. *Every bounded* $\mathbb{F}$ *martingale is a* $\mathbb{G}$ *martingale.*

3. *For every* t *and every integrable* $\mathcal{F}_\infty$ *measurable* X, $\mathbb{E}[X|\mathcal{F}_t] = \mathbb{E}[X|\mathcal{G}_t]$ $\mathbb{P}$*-a.s.*

4. *For every* t *and every integrable* $\mathcal{G}_t$ *measurable* X, $\mathbb{E}[X|\mathcal{F}_t] = \mathbb{E}[X|\mathcal{F}_\infty]$ $\mathbb{P}$*-a.s.*

B.2 Kolmogorov Continuity and Tightness

The following two theorems are taken from [55] on pages 57 and 313 respectively, where they are proved in sufficient generality for this thesis. The statements have been adjusted, but remain true.

Theorem B.2.1 (Kolmogorov Continuity). *Let X be a process on $\mathcal{I}$ with values in a Polish space $(\mathcal{Y}, d_{\mathcal{Y}})$ and assume that for some constants $a, b, c > 0$ and any $s, t \in \mathcal{I}$ such that $|t-s| \leq 1$*

$$\mathbb{E}[d_{\mathcal{Y}}(X_t - X_s)^a] \leq c|t-s|^{1+b}.$$

Then, X has a continuous version and for any $\gamma \in (0, b/a)$ the latter is almost surely locally γ Hölder continuous.

Theorem B.2.2. *Let $\{X^n\}$ be a family of continuous processes on $\mathcal{I}$ with values in a Polish space $(\mathcal{Y}, d_{\mathcal{Y}})$. Assume that $\{X^n_0\}$ is tight and that for some constants $a, b, c > 0$ and any $s, t \in \mathcal{I}$ such that $|t-s| \leq 1$ and uniformly in $n \in \mathbb{N}$,*

$$\mathbb{E}[d_{\mathcal{Y}}(X^n_t - X^n_s)^a] \leq c|t-s|^{1+b}.$$

Then, $\{X^n\}$ is tight in $C(\mathcal{I}, \mathcal{Y})$ and for any $\gamma \in (0, b/a)$ the limiting processes are almost surely locally γ Hölder continuous.

B.3 Lemmas B.3.1 and B.3.2

The authors expect that the following lemma has been proved elsewhere, but cannot yet find a reference.

Lemma B.3.1 (Fubini-type Theorem for Conditional Expectation and Itô Integrals). *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, three filtrations $\mathbb{F}^j := (\mathcal{F}^j_t)_{t \in \mathcal{I}}$ $j = 1, 2, 3$ and three processes B, H, W satisfying the following conditions:*

i) $\mathbb{F}^1 \subseteq \mathbb{F}^2 \subseteq \mathbb{F}^3$ *i.e.* $\forall t \in \mathcal{I}$, $\mathcal{F}^1_t \subseteq \mathcal{F}^2_t \subseteq \mathcal{F}^3_t$.

ii) $\mathbb{F}^1$ *is immersed in* $\mathbb{F}^2$ *under* $\mathbb{P}$.

iii) H *is a bounded* $\mathbb{F}^2$*-predictable process.*

iv) B *and* W *are* $\mathbb{F}^3$ *Brownian Motions.*

v) B *is* $\mathbb{F}^1$ *adapted.*

vi) *For any* $s, t \in \mathcal{I}$, $s \leq t$, $\sigma(W_r - W_s : s \leq r \leq t) \perp\!\!\!\perp \mathbb{F}^1_t \vee \mathbb{F}^2_s$.

Then the following hold $\mathbb{P}$*-a.s. for all* $t \in \mathcal{I}$*:*

$$\mathbb{E}\left[\int_0^t H_s\, dW_s \middle| \mathcal{F}^1_t\right] = 0, \tag{B.3.1}$$

$$\mathbb{E}\left[\int_0^t H_s\, dB_s \middle| \mathcal{F}^1_t\right] = \int_0^t \mathbb{E}[H_s|\mathcal{F}^1_s]\, dB_s. \tag{B.3.2}$$

Proof of Lemma B.3.1. The proof will follow a monotone class argument. Firstly, equations (B.3.1) and (B.3.2) are shown to hold for the family of simple predictable processes.

Let H^n be a simple predictable process defined by

$$H_t^n := Z^0 \mathbb{1}_{\{0\}}(t) + \sum_{i=0}^{n-1} Z^i \mathbb{1}_{(t_i, t_{i+1}]}(t)$$

where $n \in \mathbb{N}$, $0 \leq t_0 \leq \cdots \leq t_i \leq \cdots \leq t_n < \infty$ and Z^i are bounded $\mathcal{F}^2_{t_i}$ measurable random elements for all $i = 0, .., n$. Then (B.3.1) is verified via the following:

$$\begin{aligned}
\mathbb{E}\left[\int_0^t H_s^n \, dW_s \middle| \mathcal{F}_t^1 \right] &= \sum_{i=0}^{n-1} \mathbb{E}[Z^i (W_{t_{i+1} \wedge t} - W_{t_i}) | \mathcal{F}_t^1] \\
&= \sum_{i=0}^{n-1} \mathbb{E}[\mathbb{E}[Z^i (W_{t_{i+1} \wedge t} - W_{t_i}) | \mathcal{F}_t^1 \vee \mathcal{F}_{t_i}^2] | \mathcal{F}_t^1] \\
&= \sum_{i=0}^{n-1} \mathbb{E}[\mathbb{E}[(W_{t_{i+1} \wedge t} - W_{t_i}) | \mathcal{F}_t^1 \vee \mathcal{F}_{t_i}^2] Z^i | \mathcal{F}_t^1] \\
&= 0.
\end{aligned}$$

The first equality follows from H^n being a simple predictable process, the second and third from the tower and pull out properties of conditional expectation respectively, the fourth from condition $iv)$ and $vi)$.

To verify the second equation (B.3.2), consider the following equalities:

$$\begin{aligned}
\mathbb{E}\left[\int_0^t H_s^n \, dB_s \middle| \mathcal{F}_t^1 \right] &= \mathbb{E}\left[\sum_{i=0}^{n-1} Z^i (B_{t_{i+1} \wedge t} - B_{t_i \wedge t}) \middle| \mathcal{F}_t^1 \right] \\
&= \sum_{i=0}^{n-1} \mathbb{E}[Z^i | \mathcal{F}_t^1] (B_{t_{i+1} \wedge t} - B_{t_i \wedge t}) \\
&= \sum_{i=0}^{n-1} \mathbb{E}[Z^i | \mathcal{F}_{t_i}^1] (B_{t_{i+1} \wedge t} - B_{t_i \wedge t}) \\
&= \int_0^t \mathbb{E}[H_s^n | \mathcal{F}_s^1] \, dB_s.
\end{aligned}$$

The second equality can be seen to hold by considering separately the cases: $t < t_i$, $t_i \leq t \leq t_{i+1}$ and $t_{i+1} < t$. The third equality holds from the immersion of $\mathbb{F}^1$ in $\mathbb{F}^2$ and the fourth from the definition of H^n.

Now that the desired equalities have been established for simple predictable processes, it remains to show the equality holds for a predictable process H satisfying $iii)$ with a sequence of simple predictable processes $H^n \to H$ in uniformly on compact sets in probability (in ucp) as $n \to \infty$. Note that the sequence H^n can be chosen such that for any $n \in \mathbb{N}$, $|H^n| < K$, where K is the bound for H. Recall that convergence in ucp means that for any $t \in I$, $\sup_{0 \leq s \leq t} |H_s^n - H_s|$ converges to 0 in probability. Hence there exists a subsequence n_k that elevates the convergence to almost sure convergence along this subsequence. Therefore, by application of the dominated convergence for stochastic integrals [Theorem 32 p.145 [87]](with another subsequence) and dominated convergence for conditional expectation, the lemma is proved.

□

Lemma B.3.2. *Given a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ supporting a continuous $\mathbb{R}^{dx}$ valued stochastic process X on the interval I. Suppose that for any $T < \infty$, $\mathbb{E}[\sup_{t\in I:t\leq T}|X_t|^p] < \infty$. Then for a filtration $\mathbb{F} = (\mathcal{F}_t)_{t\in I}$ there is a $\mathcal{P}_p(\mathcal{C})$ valued $\mathbb{F}$ adapted stochastic process μ such that for all $t \in I$, $\mu_t = \mathscr{L}(X_{\cdot\wedge t}|\mathcal{F}_t)_{t\in I}$ i.e. μ_t is a regular conditional distribution of $X_{\cdot\wedge t}$ given $\mathcal{F}_t$.*

Proof of Lemma B.3.2.
For each $t \in I$, use the existence theorem for regular conditional distributions to get hold of a stochastic kernel $\kappa_{X_{\cdot\wedge t},\mathcal{F}_t}$, a $(\Omega, \mathcal{F}_t) \to (\mathcal{P}(\mathcal{C}), \mathcal{B}(\mathcal{P}(\mathcal{C})))$ measurable function.

Let $D_t := \{\omega : \kappa_{X_{\cdot\wedge t},\mathcal{F}_t} \notin \mathcal{P}_p(\mathcal{C})\}$. To see that D_t is in $\mathcal{F}_t$ first note that for some fixed $\eta \in \mathcal{P}_p(\mathcal{C})$, the sets defined $A^\eta_\varepsilon := \{\nu \in \mathcal{P}_p(\mathcal{C}) : W_p(\nu, \eta) < \varepsilon\}$ for any $\varepsilon > 0$, are in $\mathcal{B}(\mathcal{P}(\mathcal{C}))$. Note that $\mathcal{P}_p(\mathcal{C}) = \cup_{\varepsilon>0}A^\eta_\varepsilon$ and so $\mathcal{P}_p(\mathcal{C}) \in \mathcal{B}(\mathcal{P}(\mathcal{C}))$. This means that $D^c_t = \{\omega : \kappa_{X_{\cdot\wedge t},\mathcal{F}_t} \in \mathcal{P}_p(\mathcal{C})\} \in \mathcal{F}_t$ by the aforementioned measurability of $\kappa_{X_{\cdot\wedge t},\mathcal{F}_t}$ and therefore D_t is also in $\mathcal{F}_t$.

Now assume for the sake of contradiction that D_t has non-zero probability under $\mathbb{P}$. Then,

$$\begin{aligned}\mathbb{E}[\sup_{0\leq s\leq t}|X_s|^p] = \mathbb{E}[\sup_{0\leq s\leq t}|X_s|^p(\mathbb{1}_{D_t} + \mathbb{1}_{D_t{}^c})] =&\mathbb{E}[\mathbb{E}[\sup_{0\leq s\leq t}|X_s|^p|\mathcal{F}_t](\mathbb{1}_{D_t} + \mathbb{1}_{D_t{}^c})] \\ =&\infty,\end{aligned}$$

which is a contradiction.

Finally, for some arbitrary but fixed distribution $\mu \in \mathcal{P}_p(\mathcal{C})$ defining for all $t \in I$, $\mathscr{L}(X_t|\mathcal{F}_t) := \kappa_{X_t,\mathcal{F}_t}\mathbb{1}_{D^c_t} + \mu\mathbb{1}_{D_t}$ see that $\mathscr{L}(X_{\cdot\wedge t}|\mathcal{F}_t)$ is an $\mathcal{F}_t$-measurable $\mathcal{P}_p(\mathcal{C})$ valued version of the regular conditional distribution of $X_{\cdot\wedge t}$ given $\mathcal{F}_t$ for each $t \in I$.

□

Bibliography

[1] S. Ahuja. "Wellposedness of Mean Field Games with Common Noise under a Weak Monotonicity Condition". In: *SIAM Journal on Control and Optimization* 54.1 (2016), pp. 30–48.

[2] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows in Metric Spaces and in the Space of Probability Measures*. Birkhauser, 2008.

[3] G. B. Arous and O. Zeitouni. "Increasing propagation of chaos for mean field models". In: *Annales de l'Institut Henri Poincaré (B) Probability and Statistics* 35 (1999), pp. 85–102.

[4] G. B. Arous and M. D. Brunaud. "Methode de laplace: etude variationnelle des fluctuations de diffusions de type". In: *Stochastics and Stochastic Reports* 31.1-4 (1990), pp. 79–144.

[5] V. Barbu and M. Röckner. "From nonlinear Fokker–Planck equations to solutions of distribution dependent SDE". In: *Ann. Probab.* 48.4 (July 2020), pp. 1902–1920.

[6] V. Barbu, M. Röckner, and F. Russo. "Doubly probabilistic representation for the stochastic porous media type equation". In: *Ann. Inst. H. Poincaré Probab. Statist.* 53.4 (2017), pp. 2043–2073.

[7] M. Beiglbock and D. Lacker. "Denseness of adapted processes among causal couplings." In: *arXiv:1805.03185* (2020).

[8] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, 1999.

[9] D. Blackwell and L. E. Dubins. "An Extension of Skorokhod's Almost Sure Representation Theorem". In: *Proceedings of the American Mathematical Society* (1983).

[10] V. I. Bogachev, N. V. Krylov, M. Röckner, and S. V. Shaposhnikov. *Fokker–Planck–Kolmogorov Equations*. AMS, 2015.

[11] V. I Bogachev, M Röckner, and S. V Shaposhnikov. "Distances between transition probabilities of diffusions and applications to nonlinear Fokker–Planck–Kolmogorov equations". In: *Journal of Functional Analysis* 271.5 (2016), pp. 1262–1300.

[12] V. I. Bogachev. *Measure Theory*. Springer, 2007.

[13] V. I. Bogachev, M. Röckner, and S. V. Shaposhnikov. "Convergence in variation of solutions of nonlinear Fokker–Planck–Kolmogorov equations to stationary measures". In: *Journal of Functional Analysis* 276.12 (2019), pp. 3681 –3713.

[14] F. Bolley, J. A. Cañizo, and J. A. Carrillo. “Stochastic mean-field limit: non-Lipschitz forces and swarming”. In: *Math. Models Methods Appl. Sci.* 21.11 (2011), pp. 2179–2210.

[15] F. Bolley, A. Guillin, and C. Villani. “Quantitative concentration inequalities for empirical measures on non-compact spaces”. In: *Probability Theory and Related Fields* 137.3-4 (2007), pp. 541–593.

[16] M. Bossy and J.-F. Jabir. “On the Wellposedness of Some McKean Models with Moderated or Singular Diffusion Coefficient”. In: *Frontiers in Stochastic Analysis–BSDEs, SPDEs and their Applications*. Cham: Springer International Publishing, 2019, pp. 43–87.

[17] M. Bossy, J.-F. Jabir, and D. Talay. “On conditional McKean Lagrangian stochastic models”. In: *Probability theory and related fields* 151.1-2 (2011), pp. 319–351.

[18] P. Briand, P. Cardaliaguet, P.-E. Chaudru de Raynal, and Y. Hu. “Forward and backward stochastic differential equations with normal constraints in law”. In: *Stochastic Processes and their Applications* 130.12 (2020), pp. 7021 –7097.

[19] R. Buckdahn, J. Li, S. Peng, and C. Rainer. “Mean-field stochastic differential equations and associated PDEs”. In: *The Annals of Probability* 45.2 (2017), pp. 824–878.

[20] L. Campi and M. Fischer. “N-player games and mean-field games with absorption”. In: *Ann. Appl. Probab.* 28.4 (2018), pp. 2188–2242.

[21] P. Cardaliaguet, F. Delarue, J. Lasry, and P. Lions. *The Master Equation and the Convergence Problem in Mean Field Games: (AMS-201)*. Annals of Mathematics Studies. Princeton University Press, 2019.

[22] P. Cardaliaguet. *Notes on Mean Field Games (from P.-L. Lions' lectures at Collège de France)*. Online at https://www.ceremade.dauphine.fr/~cardaliaguet/MFG20130420.pdf. 2013.

[23] P. Cardaliaguet and P. Souganidis. “On first order mean field game systems with a common noise”. In: *arXiv: 2009.12134* (2020).

[24] R. Carmona, F. Delarue, and D. Lacker. “Mean field games with common noise”. In: *The Annals of Probability* (2016).

[25] R. Carmona and X. Zhu. “A probabilistic approach to mean field games with major and minor players”. In: *Ann. Appl. Probab.* 26.3 (June 2016), pp. 1535–1580.

[26] R. Carmona and F. Delarue. *Probabilistic Theory of Mean Field Games with Applications I-II*. Springer, 2017.

[27] J.-F. Chassagneux, D. Crisan, and F. Delarue. “Classical solutions to the master equation for large population equilibria”. In: *arXiv:1411.3009* (2014).

[28] J.-F. Chassagneux, L. Szpruch, and A. Tse. “Weak quantitative propagation of chaos via differential calculus on the space of measures”. In: *arXiv:1901.02556* (2019).

[29] P.-E. Chaudru de Raynal. "Strong well posedness of McKean–Vlasov stochastic differential equations with Hölder drift". In: *Stochastic Processes and their Applications* 130.1 (2020), pp. 79 –107.

[30] P.-E. Chaudru de Raynal and N. Frikha. "Well-posedness for some non-linear diffusion processes and related PDE on the Wasserstein space". In: *arXiv:1811.06904* (2018).

[31] M. Coghi and B. Gess. "Stochastic nonlinear Fokker-Planck equations". In: *Nonlinear Analysis* 187 (2019).

[32] D. Crisan, C. Janjigian, and T. G. Kurtz. "Particle representations for stochastic partial differential equations with boundary conditions". In: *Electron. J. Probab.* 23 (2018), 29 pp.

[33] J. Dawson D. & Vaillancourt. "Stochastic McKean-Vlasov equations". In: *J. NoDEA* (1995).

[34] F. Delarue. "Restoring Uniqueness to Mean-Field Games by Randomizing the Equilibria". In: *Stochastics and Partial Differential Equations: Analysis and Computations* (2019).

[35] F. Delarue and R. Foguen Tchuendom. "Selection of equilibria in a linear quadratic mean-field game". In: *Stochastic Processes and their Applications* (2019).

[36] R. M. Dudley. *Real Analysis and Probability*. Cambridge Univ Press, 2002.

[37] A. Figalli. "Existence and uniqueness of martingale solutions for SDEs with rough or degenerate coefficients". In: *Journal of Functional Analysis* 254.1 (2008), pp. 109 –153.

[38] N. Fournier and B. Jourdain. "Stochastic particle approximation of the Keller–Segel equation and two-dimensional generalization of Bessel processes". In: *Ann. Appl. Probab.* 27.5 (2017), pp. 2807–2861.

[39] T. Funaki. "A certain class of diffusion processes associated with nonlinear parabolic equations". In: *Probability Theory and Related Fields* 67.3 (1984), pp. 331–348.

[40] J. Gärtner. "On the McKean-Vlasov Limit for Interacting Diffusions". In: *Mathematische Nachrichten* 137.1 (1988), pp. 197–248.

[41] I. Gyöngy and N. Krylov. "Existence of strong solutions for Itô's stochastic equations via approximations". In: *Probability Theory and Related Fields* 105.2 (1996), pp. 143–158.

[42] B. Hambly and A. Søjmark. "An SPDE model for systemic risk with endogenous contagion". In: *Finance and Stochastics* 23.3 (2019), pp. 535–594.

[43] P. R. Hamlos. *Measure Theory*. Van Nostrand, 1950.

[44] W. R. P. Hammersley, D. Šiška, and Ł. Szpruch. "McKean-Vlasov SDEs under Measure Dependent Lyapunov Conditions". In: *arXiv:1802.03974* (2018).

[45] W. R. P. Hammersley, D. Šiška, and Ł. Szpruch. "Weak Existence and Uniqueness for McKean-Vlasov SDEs with Common Noise". In: *arXiv:1908.00955* (2019).

[46] M. Huang. "Large-Population LQG Games Involving a Major Player: The Nash Certainty Equivalence Principle". In: *SIAM Journal on Control and Optimization* 48.5 (2010), pp. 3318–3353.

[47] M. Huang, R. P. Malhamé, and P. E. Caines. "Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle". In: *Commun. Inf. Syst.* 6.3 (2006), pp. 221–252.

[48] M. Huang, R. P. Malhamé, and P. E. Caines. "The Nash certainty equivalence principle and McKean-Vlasov systems: An invariance principle and entry adaptation". In: *2007 46th IEEE Conference on Decision and Control*. 2007, pp. 121–126.

[49] X. Huang. "Path-Distribution Dependent SDEs with Singular Coefficients". In: *arXiv:1902.08953* (2019).

[50] X. Huang and F.-Y. Wang. "Distribution dependent SDEs with singular coefficients". In: *Stochastic Processes and their Applications* (2018).

[51] P.-E. Jabin and Z. Wang. "Quantitative estimates of propagation of chaos for stochastic systems with $W^{(-1,\infty)}$ kernels". In: *Inventiones mathematicae* 214.1 (2018), pp. 523–591.

[52] J.-F. Jabir. "Rate of propagation of chaos for diffusive stochastic particle systems via Girsanov transformation". In: *arXiv:1907.09096* (2019).

[53] J. Jacod and J. Mémin. "Sur un type de convergence intermédiaire entre la convergence en loi et la convergence en probabilité". In: *Séminaire de Probabilités XV*. Springer, 1981.

[54] M. Kac. "Foundations of kinetic theory". In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. III*. University of California Press, Berkeley and Los Angeles, 1956, pp. 171–197.

[55] O. Kallenberg. *Foundations of Modern Probability*. Springer, 2002.

[56] O. Kallenberg. *Random Measures, Theory and Applications*. 1st. 77 vols. Probability Theory and Stochastic Modelling. Springer International Publishing, 2017. 680 pp.

[57] I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*. Springer, 2012.

[58] R. Khasminskii. *Stochastic stability of differential equations*. Vol. 7. Monographs and Textbooks on Mechanics of Solids and Fluids: Mechanics and Analysis. Translated from the Russian by D. Louvish. Sijthoff & Noordhoff, Alphen aan den Rijn—Germantown, Md., 1980, pp. xvi+344.

[59] V. Kolokoltsov and M. Troeva. "Regularity and Sensitivity for McKean-Vlasov Type SPDEs Generated by Stable-like Processes". In: *Issues of Analysis* 25 (2018), pp. 69–81.

[60] V. N. Kolokoltsov and M. Troeva. "On mean field games with common noise and McKean-Vlasov SPDEs". In: *Stochastic Analysis and Applications* 37.4 (2019), pp. 522–549.

[61] V. N. Kolokoltsov and M. Troeva. "On Mean Field Games with Common Noise based on Stable-Like Processes". English. In: *Statistics, Optimization & Information Computing* 7.2 (2019), pp. 264–276.

[62] V. N. Kolokoltsov and M. Troeva. "Regularity and sensitivity for McKean-Vlasov SPDEs". In: *AIP Conference Proceedings* 1907.1 (2017), p. 030046.

[63] N. Kryloff and N. Bogoliouboff. "La Theorie Generale De La Mesure Dans Son Application A L'Etude Des Systemes Dynamiques De la Mecanique Non Lineaire". In: *Annals of Mathematics* 38.1 (1937), pp. 65–113.

[64] N. V. Krylov. *Controlled diffusion processes*. Vol. 14. Applications of Mathematics. Translated from the Russian by A. B. Aries. Springer-Verlag, New York-Berlin, 1980, pp. xii+308.

[65] T. Kurtz. "Weak and strong solutions of general stochastic models". In: *Electron. Commun. Probab.* 19 (2014).

[66] T. G. Kurtz and J. Xiong. "Numerical Solutions for a Class of SPDEs with Application to Filtering". In: *Stochastics in Finite and Infinite Dimensions: In Honor of Gopinath Kallianpur*. Birkhäuser Boston, 2001, pp. 233–258.

[67] T. G. Kurtz and J. Xiong. "Particle representations for a class of nonlinear SPDEs". In: *Stochastic Processes and their Applications* (1999).

[68] D. Lacker. "A general characterization of the mean field limit for stochastic differential games". In: *Probability Theory and Related Fields* 165 (2014).

[69] D. Lacker. "Dense sets of joint distributions appearing in filtration enlargements, stochastic control, and causal optimal transport". In: *arXiv:1805.03185* (2018).

[70] D. Lacker. "On a strong form of propagation of chaos for McKean-Vlasov equations". In: *Electronic Communications in Probability* (2018).

[71] D. Lacker, M. Shkolnikov, and J. Zhang. "Superposition and mimicking theorems for conditional McKean-Vlasov equations". In: *arXiv:2004:00099* (2020).

[72] J.-M. Lasry and P.-L. Lions. "Jeux à champ moyen. I – Le cas stationnaire". In: *Comptes Rendus Mathematique* 343.9 (2006), pp. 619 –625.

[73] J.-M. Lasry and P.-L. Lions. "Jeux à champ moyen. II – Horizon fini et contrôle optimal". In: *Comptes Rendus Mathematique* 343.10 (2006), pp. 679 –684.

[74] J.-M. Lasry and P.-L. Lions. "Mean field games". In: *Japanese Journal of Mathematics* 2.1 (2007), pp. 229–260.

[75] S. J. Ledger and A. Søjmark. "At the Mercy of the Common Noise: Blow-ups in a Conditional McKean–Vlasov Problem". In: *arXiv:1807.05126* (2018).

[76] S. Ledger and A. Søjmark. "Uniqueness for contagious McKean–Vlasov systems in the weak feedback regime". In: *Bulletin of the London Mathematical Society* 52.3 (2020), pp. 448–463.

[77] J. Li and H. Min. "Weak solutions of mean-field stochastic differential equations". In: *Stochastic Analysis and Applications* 35.3 (2017), pp. 542–568.

[78] P.-L. Lions. "Théorie des jeux à champs moyen et applications." In: *Lectures at the Collège de France.* (2007).

[79] E. Luçon and W. Stannat. "Mean field limit for disordered diffusions with singular interactions". In: *Ann. Appl. Probab.* 24.5 (2014), pp. 1946–1993.

[80] H. P. McKean Jr. "A class of Markov processes associated with nonlinear parabolic equations". In: *Proceedings of the National Academy of Sciences of the United States of America* 56.6 (1966), p. 1907.

[81] H. P. McKean Jr. "Propagation of chaos for a class of nonlinear parabolic equations". In: *Lecture series in differential equations, Catholic University, Washington DC* 7 (1967), pp. 41–57.

[82] S. Mehri and W. Stannat. "Weak solutions to Vlasov–McKean equations under Lyapunov-type conditions". In: *Stochastics and Dynamics* (2019).

[83] S. Méléard. "Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models". In: *Probabilistic models for nonlinear partial differential equations*. Springer, 1996, pp. 42–95.

[84] Y. S. Mishura and A. Y. Veretennikov. "Existence and uniqueness theorems for solutions of McKean–Vlasov stochastic equations". In: *arXiv:1603.02212* (2016).

[85] K. Oelschlager. "A Martingale Approach to the Law of Large Numbers for Weakly Interacting Stochastic Processes". In: *Ann. Probab.* 12.2 (May 1984), pp. 458–479.

[86] G. D. Prato. *An Introduction to Infinite-Dimensional Analysis*. Universitext. Springer-Verlag Berlin Heidelberg, 2006.

[87] P. Protter. *Stochastic Integration and Differential Equations*. Springer, 1990.

[88] M. Röckner and X. Zhang. "Well-posedness of distribution dependent SDEs with singular drifts". In: *arXiv:1809.02216* (2018).

[89] G. dos Reis, W. Salkeld, and J. Tugaut. "Freidlin–Wentzell LDP in path space for McKean–Vlasov equations and the functional iterated logarithm law". In: *Ann. Appl. Probab.* 29.3 (June 2019), pp. 1487–1540.

[90] M. Scheutzow. "Uniqueness and nonuniqueness of solutions of Vlasov–McKean equations". In: *J. Austral. Math. Soc. Ser. A* 43.2 (1987), pp. 246–256.

[91] A. V. Skorokhod. *Studies in the theory of random processes*. Addison–Wesley, Reading, Mass., 1965.

[92] D. W. Stroock and S. R. S. Varadhan. *Multidimensional diffusion processes*. Classics in Mathematics. Reprint of the 1997 edition. Springer-Verlag, Berlin, 2006, pp. xii+338.

[93] A.-S. Sznitman. *Topics in propagation of chaos*. Springer, 1991, pp. 165–251.

[94] Ł. Szpruch and X. Zhāng. "V-integrability, asymptotic stability and comparison property of explicit numerical schemes for non-linear SDEs". In: *Math. Comp.* 87.310 (2018), pp. 755–783.

[95] G. Taraldsen. "Optimal Learning from the Doob-Dynkin lemma". In: *arXiv:1801.00974* (2018).

[96] R. F. Tchuendom. "Uniqueness for Linear-Quadratic Mean Field Games with Common Noise". In: *Dynamic Games and Applications* 8.1 (2018), pp. 199–210.

[97] D. Trevisan. "Well-posedness of multidimensional diffusion processes with weakly differentiable coefficients". In: *Electron. J. Probab.* 21 (2016), 41 pp.

[98] J. Vaillancourt. "On the existence of random mckean–vlasov limits for triangular arrays of exchangeable diffusions". In: *Stochastic Analysis and Applications* 6.4 (1988), pp. 431–446.

[99] S. S. Vallender. "Calculation of the Wasserstein Distance Between Probability Distributions on the Line". In: *Theory Probab. Appl.* 18.4 (1972).

[100] C. Villani. *Optimal Transport: Old and New*. Springer, 2009.

[101] F.-Y. Wang. "Distribution dependent SDEs for Landau type equations". In: *Stochastic Processes and their Applications* 128.2 (2018), pp. 595 –621.