

THE UNIVERSITY OF CHICAGO

HOMOGENIZATION AND CALIBRATION PROBLEMS THROUGH THE LENS OF
MEAN FIELD GAMES AND MEAN FIELD CONTROL

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF MATHEMATICS

BY
ANTONIOS ZITRIDIS

CHICAGO, ILLINOIS
AUGUST 2025

Copyright © 2025 by Antonios Zitridis
All Rights Reserved

To my family, mentors and friends,

TABLE OF CONTENTS

ACKNOWLEDGMENTS	vi
ABSTRACT	vii
1 INTRODUCTION	1
1.1 Homogenization of slow-fast systems of SDEs	1
1.2 Singular perturbations of Hamilton-Jacobi equations in the Wasserstein space	3
1.3 Model calibration via martingale optimal transport and martingale Schrödinger bridges	6
2 HOMOGENIZATION OF SLOW-FAST CONDITIONAL MCKEAN-VLASOV SDES	11
2.1 Introduction	11
2.1.1 Notation	14
2.1.2 The main result	15
2.1.3 Summary of the chapter	18
2.2 Wasserstein space and Wasserstein differentiation	18
2.3 Assumptions and Preliminary results	22
2.3.1 Assumptions	25
2.3.2 Preliminary results	27
2.4 The Poisson Equation	28
2.5 Cauchy problem associated to McKean-Vlasov processes	33
2.6 Proof of the main result	34
2.6.1 Heuristic calculation	34
2.6.2 Proof of the main theorem	35
2.7 On the McKean-Vlasov control equation	40
2.8 On the choice of the Assumptions	52
3 SINGULAR PERTURBATION OF HAMILTON-JACOBI-BELLMAN EQUATIONS IN THE WASSERSTEIN SPACE	55
3.1 Introduction	55
3.1.1 Background	55
3.1.2 Slow-fast systems and slow-fast stochastic optimal control problems	58
3.1.3 Extensions and Outlook	61
3.1.4 Organization of the chapter.	62
3.2 Notation and Assumptions	63
3.3 On viscosity solutions	65
3.4 Proof of the main result	70
3.4.1 Formal calculation	70
3.4.2 Proof of Theorem 3.1.1	73
3.5 Wasserstein space and Wasserstein Derivatives	88
3.6 Technical proofs of the chapter	90

4	CALIBRATION OF STOCHASTIC VOLATILITY MODELS VIA MARTINGALE OPTIMAL TRANSPORT AND MARTINGALE SCHRÖDINGER BRIDGES . .	103
4.1	Introduction	103
4.1.1	Overview	104
4.1.2	Calibration via martingale optimal transport	105
4.1.3	Calibration via martingale Schrödinger bridges	106
4.1.4	Main results	108
4.1.5	Organization of the chapter	112
4.2	Notation, Assumptions and Preliminary results	112
4.2.1	Notation	112
4.2.2	Assumptions	113
4.2.3	Preliminary results	114
4.3	Hamilton-Jacobi Equations	116
4.3.1	A regularity result for a special case	119
4.4	Martingale Optimal Transport, Proof of Theorem 4.1.1	120
4.5	Martingale Schrödinger bridges, Theorems 4.1.3, 4.1.4	130
4.6	Important results from the literature	142
	REFERENCES	144

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest gratitude to my advisor, Professor Panagiotis Souganidis. His guidance, insight, and unwavering support have shaped my mathematical thinking and profoundly influenced my academic development.

I would also like to thank Professor Luis Silvestre for his support, his teaching, and for serving on my Ph.D. thesis committee.

I am deeply grateful to the members of my research group: Nikiforos Mimikos, Sebastian Muñoz, Peter Morfe, Mark Cerenzia and Joe Jackson. Our discussions, collaborations, and shared experiences have been an essential part of my journey.

To the broader mathematics community at the University of Chicago, thank you for creating such a vibrant, supportive, and inspiring environment. In particular, I want to thank Jinwoo Sung, Daniel Mitsutani, Adrian Chu, Nikolay Grantcharov, DeVon Ingram, Xinchun Ma, David Bowman, Seyhun Ji, Kapil Chandran, Aaron Slipper, Andres Contreras Hip, Kostas Psaromiligkos, Faidon Andriopoulos, Jason Kountouridis and the postdoc Sayan Das for making my time here truly special.

I am also grateful to my friends from Greece: Stavros, Gerasimos, Nikos, Minas, Kostis and Nasos for being such wonderful company, even from the other side of the world most of the time.

Finally, I want to thank my parents, my brother and Alicia for their unconditional love and support throughout this journey.

ABSTRACT

In this thesis, we apply techniques from mean field game theory and mean field control theory to establish homogenization results for various types of slow-fast stochastic differential equations (SDEs) and to address calibration problems for stochastic volatility models.

The first part is devoted to proving a homogenization result for conditional slow-fast McKean–Vlasov SDEs. In the second part, we extend this analysis to the controlled setting, establishing a homogenization result for Hamilton–Jacobi equations posed in the Wasserstein space. The final part focuses on the use of martingale optimal transport and martingale Schrödinger bridges as tools for calibrating stochastic volatility models, either to SPX option data or jointly to SPX and VIX options.

CHAPTER 1

INTRODUCTION

The goal of this thesis is to apply techniques from the recently developed mean field game theory and, its twin, mean field control theory, to study the homogenization of various types of slow-fast McKean-Vlasov stochastic differential equations (SDEs) and to address calibration problems for stochastic volatility models via martingale optimal transport and martingale Schrödinger bridges. We give an overview of the problems and the main results.

1.1 Homogenization of slow-fast systems of SDEs

Slow-fast systems are commonly observed in nature, such as in weather modeling, where rapid atmospheric changes interact with the more slowly evolving ocean, as well as in population dynamics and celestial mechanics. These systems are typically represented by a pair of stochastic processes $(X_t^\varepsilon, Y_t^\varepsilon)$ for $\varepsilon > 0$, where X_t^ε represents the slow variable and Y_t^ε the fast variable. An important question is to characterize their behavior as ε tends to 0. To formalize the problem, we let $(\Omega^i, \mathbb{F}^i = (\mathcal{F}_t^i)_{t \geq 0}, \mathbb{P}^i)$, $i = 0, 1$ be two filtered probability spaces, B_t, W_t two independent standard n -dimensional $\mathbb{F}^1$ -Brownian motions, and B_t^0 an n -dimensional $\mathbb{F}^0$ -Brownian motion representing the common noise. For a parameter $\lambda \in \mathbb{R}$, consider the system of conditional slow-fast McKean-Vlasov SDEs:

$$\begin{aligned} dX_t^\varepsilon &= \left(\frac{\lambda}{\varepsilon} f + c \right) (X_t^\varepsilon, Y_t^\varepsilon, \mathcal{L}^0(X_t^\varepsilon)) dt + \sigma(X_t^\varepsilon, Y_t^\varepsilon, \mathcal{L}^0(X_t^\varepsilon)) dW_t + \sigma_1^0(X_t^\varepsilon, \mathcal{L}^0(X_t^\varepsilon)) dB_t^0, \\ dY_t^\varepsilon &= \frac{1}{\varepsilon} \left(\frac{1}{\varepsilon} f(X_t^\varepsilon, Y_t^\varepsilon, \mathcal{L}^0(X_t^\varepsilon)) + g(X_t^\varepsilon, Y_t^\varepsilon, \mathcal{L}^0(X_t^\varepsilon)) \right) dt \\ &\quad + \frac{1}{\varepsilon} \left(\tau_1(X_t^\varepsilon, Y_t^\varepsilon, \mathcal{L}^0(X_t^\varepsilon)) dW_t + \tau_2(X_t^\varepsilon, Y_t^\varepsilon, \mathcal{L}^0(X_t^\varepsilon)) dB_t + \sigma_2^0(X_t^\varepsilon, \mathcal{L}^0(X_t^\varepsilon)) dB_t^0 \right), \end{aligned} \tag{1.1.1}$$

where $\mathcal{L}^0(X_t^\varepsilon)$ denotes the conditional law of X_t^ε upon $\mathcal{F}_t^0$. The functions $b, c, f, g : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^d$, and $\sigma, \tau_1, \tau_2 : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^{d \times n}$, with $\sigma_1^0, \sigma_2^0 : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^{d \times n}$

¹. A natural question to ask is whether there exists a process $\overline{X}_t$ such that X_t^ε converges to $\overline{X}_t$ as $\varepsilon \rightarrow 0$, in some appropriate sense. In the case where $\sigma_1^0 = \sigma_2^0 = 0$, which means that the common noise B_t^0 has no impact on the system, convergence results have been established in various settings starting from Pardoux and Veretenikov [PV01b]. These results were later extended by various authors, for instance in [BS23a, RSX19a, LWX24], where the coefficients depend on the distribution of X_t and/or Y_t .

In Chapter 2, we study the natural follow-up question of what happens when $\sigma_1^0 \neq 0$ or $\sigma_2^0 \neq 0$, which corresponds to slow-fast systems with extra random factors. Our strategy relies on the analysis of the infinite dimensional partial differential equation (PDE) satisfied by $u^\varepsilon : [0, T] \times \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d) \rightarrow \mathbb{R}$ with $u^\varepsilon(s, m) = \mathbb{E}^{\mathbb{P}^0}[G(\mathcal{L}^0(X_T^{\varepsilon, m, s}))]$, where $T > 0$, $G : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ is a given function and $X_t^{\varepsilon, m, s}$ is the first component of the solution of (1.1.1) with initial condition at time $t = s$ such that $\mathcal{L}(X_s^{\varepsilon, m, s}, Y_s^{\varepsilon, m, s}) = m \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$. More specifically, under specific assumptions on the coefficients of (1.1.1) and G , we show that u^ε satisfies the following PDE in the classical sense

$$\begin{cases} \partial_s u^\varepsilon(s, m) + \mathcal{L}^\varepsilon u^\varepsilon(s, m) = 0, & \text{in } [0, T] \times \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d), \\ u^\varepsilon(T, m) = G(\pi_*^1 m), & \text{in } \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d), \end{cases} \quad (1.1.2)$$

where $\mathcal{L}^\varepsilon = \mathcal{L}_2 + \frac{1}{\varepsilon}\mathcal{L}_2 + \frac{1}{\varepsilon^2}\mathcal{L}_0$, with $\mathcal{L}_0, \mathcal{L}_1, \mathcal{L}_2$ linear operators involving derivatives with respect to the measure variable ² and $\pi_*^1 m \in \mathcal{P}_2(\mathbb{R}^d)$ is the measure such that $\pi_*^1 m(A) = m(A \times \mathbb{R}^d)$ for any measurable set $A \subset \mathbb{R}^d$. This allows us to characterize the limit of u^ε as $\varepsilon \rightarrow 0$, as stated in the following theorem.

Theorem 1.1.1. *Suppose that $G : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ is a C^4 function with bounded derivatives,*

1. Here, $\mathcal{P}_2(X)$ denotes the space of probability measures over X with finite second moments.

2. Here for derivatives with respect to the measure arguments we use Wasserstein derivatives: $D_\mu U(\mu, x) = D_x \frac{\delta U}{\delta \mu}(\mu, x)$, where $\frac{\delta U}{\delta \mu}(\mu, x)$ is such that $\lim_{t \rightarrow 0} \frac{U(t\mu + (1-t)\mu') - U(\mu)}{t} = \int \frac{\delta U}{\delta \mu}(\mu, x) d(\mu - \mu')(x)$ for every $\mu', \mu \in \mathcal{P}(\mathbb{R}^d)$.

$f, c, \sigma, g, \tau_1, \tau_2, \sigma_1^0, \sigma_2^0$ are all C^4 functions whose derivatives grow polynomially in y , f satisfies $f(x, y, \mu) \cdot y \leq C(1 - |y|^2)$ for every $(x, y, \mu) \in \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$, and the matrix $a(x, y, \mu) := \frac{1}{2}(\tau_1 \tau_1^\top + \tau_2 \tau_2^\top + \sigma_2^0 \sigma_2^{0\top})(x, y, \mu)$ is uniformly elliptic and bounded, then there exists a unique classical solution of (1.1.2). Furthermore, there exist $\bar{\gamma} : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^d$ and $\bar{D}, \Sigma : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^{d \times n}$ such that $u^\varepsilon(s, m) \xrightarrow{\varepsilon \rightarrow 0} \mathbb{E}^{\bar{\mathbb{P}}^0} \left[G \left(\mathcal{L}(\bar{X}_T^{s, \mu} | \bar{\mathcal{F}}_T^0) \right) \right]$ pointwise, where $\bar{X}_t^{s, \mu}$ satisfies $\mathcal{L}(\bar{X}_s^{s, \mu}) = \mu$ and for $t \geq s$

$$\begin{aligned} d\bar{X}_t^{s, \mu} = & \bar{\gamma} \left(\bar{X}_t^{s, \mu}, \mathcal{L}(\bar{X}_t^{s, \mu} | \bar{\mathcal{F}}_t^0) \right) dt + \bar{D} \left(\bar{X}_t^{s, \mu}, \mathcal{L}(\bar{X}_t^{s, \mu} | \bar{\mathcal{F}}_t^0) \right) d\bar{B}_t^1 \\ & + \Sigma \left(\bar{X}_t^{s, \mu}, \mathcal{L}(\bar{X}_t^{s, \mu} | \bar{\mathcal{F}}_t^0) \right) d\bar{B}_t^0, \end{aligned} \quad (1.1.3)$$

for a possibly different pair of filtered probability spaces $(\bar{\Omega}^i, \bar{\mathbb{F}}^i = (\bar{\mathcal{F}}_t^i)_{t \geq 0}, \bar{\mathbb{P}}^i)$, $i = 0, 1$ and Brownian motions $\bar{B}^0, \bar{B}^1$.

1.2 Singular perturbations of Hamilton-Jacobi equations in the Wasserstein space

Let $\varepsilon, T > 0$, $d \in \mathbb{N}^*$, A a compact subset of $\mathbb{R}^{2d}$ and $(B^k)_{k=0,1,2}$ independent d -dimensional Brownian motions defined on a fixed filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$. In Chapter 3 we consider the following controlled version of (1.1.1) that is discussed in chapter 2 on the d -dimensional torus $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$

$$\begin{cases} dX_t^\varepsilon &= c(X_t^\varepsilon, Y_t^\varepsilon, \alpha_t, \mathcal{L}^0(X_t^\varepsilon))dt + \sqrt{2}\sigma_1 dB_t^1 + \sqrt{2}\sigma_{0,1} dB_t^0, \\ dY_t^\varepsilon &= \frac{1}{\varepsilon} f(X_t^\varepsilon, Y_t^\varepsilon, \alpha_t, \mathcal{L}^0(X_t^\varepsilon))dt + \frac{\sqrt{2}}{\sqrt{\varepsilon}} \sigma_2 dB_t^2 + \frac{\sqrt{2}}{\sqrt{\varepsilon}} \sigma_{0,2} dB_t^0, \end{cases} \quad (1.2.1)$$

where $(\mathcal{F}_t^0)_{0 \leq t \leq T}$ is the filtration generated by B^0 , $\mathcal{L}^0(X_t)$ is the law of X_t conditioned upon $\mathcal{F}_t^0$, α_t is an adapted process with values in A and where $f, c : \mathbb{T}^d \times \mathbb{T}^d \times A \rightarrow \mathbb{R}^d$, $\sigma_1, \sigma_2, \sigma_{0,1}, \sigma_{0,2}$ are $d \times d$ matrices. Here B^1, B^2 play the role of the idiosyncratic noise,

while B^0 is the common noise.

We also introduce the functions $\mathcal{G} : \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}^3$, $L : \mathbb{T}^d \times \mathbb{T}^d \times A \times \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$ and we let $(t_0, m_0) \in [0, T] \times \mathcal{P}(\mathbb{T}^{2d})$. We study the behavior as $\varepsilon \rightarrow 0$ of the mean field control problem

$$U^\varepsilon(t_0, m_0) := \inf_{\alpha \in \mathcal{A}} \mathbb{E} \left[\int_{t_0}^T L \left(X_t^{t_0, \varepsilon, m_0}, Y_t^{t_0, \varepsilon, m_0}, \alpha_t, \mathcal{L}^0(X_t^{t_0, \varepsilon, m_0}) \right) dt + \mathcal{G} \left(\mathcal{L}^0(X_T^{t_0, \varepsilon, m_0}) \right) \right], \quad (1.2.2)$$

where the infimum is taken over $\mathcal{A}$ the set of all controls, i.e. the set of all $\mathbb{F}$ -adapted processes α with values in the compact set A and $(X_t^{t_0, \varepsilon, m_0}, Y_t^{t_0, \varepsilon, m_0})$ satisfies (1.2.1) with initial condition on the law $\mathcal{L}((X_{t_0}^{t_0, \varepsilon, m_0}, Y_{t_0}^{t_0, \varepsilon, m_0})) = m_0 \in \mathcal{P}(\mathbb{T}^{2d})$.

In the framework where there are no controls from the previous chapter, a crucial tool in analyzing the behavior of $\mathcal{L}^0(X_t^\varepsilon)$ is the partial differential equation (PDE) satisfied by U^ε from (1.2.2) with $L = 0$, which is $\mathcal{G}(\mathcal{L}^0(X_T^{t_0, \varepsilon, m_0})) =: V^\varepsilon(t_0, m_0)$. This was a linear PDE over the Wasserstein space $\mathcal{P}(\mathbb{T}^{2d})$ and, under suitable regularity assumptions on $c, f, \mathcal{G}$, it was shown that V^ε is its unique classical solution. In the case where c, f in (1.2.1) depend on the controls $\alpha \in \mathcal{A}$, the PDE satisfied by U^ε is no longer linear and is the Hamilton-Jacobi equation posed in $[0, T] \times \mathcal{P}(\mathbb{T}^{2d})$ ⁴

$$\begin{cases} -\partial_t U^\varepsilon + \int_{\mathbb{T}^{2d}} H^\varepsilon(x, y, D_m U^\varepsilon(t, m, x, y), \pi_*^1 m) m(dx, dy) \\ - \int_{\mathbb{T}^{2d}} \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_0^\varepsilon \sigma_0^{\varepsilon \top} D_{mm}^2 U^\varepsilon(t, m, x, y, x', y') \right) m(dx, dy) m(dx', dy') = 0, \\ - \int_{\mathbb{T}^{2d}} \text{tr} \left((\sigma^\varepsilon \sigma^{\varepsilon \top} + \sigma_0^\varepsilon \sigma_0^{\varepsilon \top}) D_{(x, y)} D_m U^\varepsilon(t, m, x, y) \right) m(dx, dy) \\ U^\varepsilon(T, m) = \mathcal{G}(\pi_*^1 m), \end{cases} \quad (1.2.3)$$

3. We use the notation $\mathcal{P}(X)$ for the Wasserstein space, that is the space of probability measures over a Polish space X equipped with the Wasserstein distance $\mathbf{d}_1$.

4. $D_m U$ denotes the Wasserstein derivative.

where for $m \in \mathcal{P}(\mathbb{T}^{2d})$, $\pi_*^1 m \in \mathcal{P}(\mathbb{T}^d)$ is the x -marginal of m ⁵, $H : \mathbb{T}^d \times \mathbb{T}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$ is the Hamiltonian ⁶

$$H(x, y, p, q, \mu) = \sup_{\alpha \in A} \{-p \cdot c(x, y, \alpha, \mu) - q \cdot f(x, y, \alpha, \mu) - L(x, y, \alpha, \mu)\} \quad (1.2.4)$$

and, for $x, y \in \mathbb{T}^d$, $p, q \in \mathbb{R}^d$, $\mu \in \mathcal{P}(\mathbb{T}^d)$

$$H^\varepsilon(x, y, p, q, \mu) := H\left(x, y, p, \frac{q}{\varepsilon}, \mu\right), \quad \sigma^\varepsilon := \begin{pmatrix} \sigma_1 & 0 \\ 0 & \frac{1}{\sqrt{\varepsilon}}\sigma_2 \end{pmatrix} \quad \text{and} \quad \sigma_0^\varepsilon := \begin{pmatrix} \sigma_{0,1} & 0 \\ \frac{1}{\sqrt{\varepsilon}}\sigma_{0,2} & 0 \end{pmatrix},$$

with 0 being the $d \times d$ zero matrix.

It is important to note that even with smooth data, infinite dimensional equations like (1.2.3) might not admit a classical solution (see [DS24, Example 2.4]), highlighting the necessity of employing weaker notions of solutions, such as viscosity solutions. In [DJS23] (among others) it was shown that under appropriate assumptions for $H, \mathcal{G}$, (1.2.3) has a unique viscosity solution. By using this well-posedness result, we prove the following theorem under appropriate assumptions for $H, \mathcal{G}, f$ and c

Theorem 1.2.1. *There exists a function $\overline{H} : \mathbb{T}^d \times \mathbb{R}^d \times \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$ such that for $(x, p, \mu) \in \mathbb{T}^d \times \mathbb{R}^d \times \mathcal{P}(\mathbb{T}^d)$, $\overline{H}(x, p, \mu)$ is the unique constant such that*

$$-tr\left((\sigma_2\sigma_2^\top + \sigma_{0,2}\sigma_{0,2}^\top)D_{yy}^2v\right) + H(x, y, p, D_yv, \mu) = \overline{H}(x, p, \mu), \quad y \in \mathbb{T}^d,$$

has a unique, up to a constant, classical solution $v(y) = v(y; x, p, \mu)$ and the equation on

5. This means that $\pi_*^1 m(A) = m(A \times \mathbb{T}^d)$, for any measurable $A \subset \mathbb{T}^d$.

6. We denote by $a \cdot b$ the dot product of $a, b \in \mathbb{R}^d$.

$[0, T] \times \mathcal{P}(\mathbb{T}^d)$

$$\left\{ \begin{array}{l} -\partial_t U + \int_{\mathbb{T}^d} \overline{H}(x, D_\mu U(t, \mu, x), \mu) \mu(dx) \\ \quad - \int_{\mathbb{T}^d} \text{tr} \left((\sigma_1 \sigma_1^\top + \sigma_{0,1} \sigma_{0,1}^\top) D_x D_\mu U(t, \mu, x) \right) \mu(dx) \\ \quad - \int_{\mathbb{T}^d} \int_{\mathbb{T}^d} \text{tr} \left(\sigma_{0,1} \sigma_{0,1}^\top D_{\mu\mu}^2 U(t, \mu, x, y) \right) \mu(dx) \mu(dy) = 0, \\ U(T, \mu) = \mathcal{G}(\mu), \end{array} \right. \quad (1.2.5)$$

has a unique viscosity solution $U : [0, T] \times \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$. Furthermore, $U^\varepsilon(t_0, m_0)$ converges as $\varepsilon \rightarrow 0$ to $U(t_0, \pi_*^1 m_0)$ uniformly in $[0, T] \times \mathcal{P}(\mathbb{T}^{2d})$.

The above theorem extends the result of Bensoussan and Blankenship [BB05] for Hamilton-Jacobi equations in finite dimensions to equations on the Wasserstein space. In the literature, $\overline{H}$ is called effective Hamiltonian.

The proof of Theorem 1.2.1 requires that $\overline{H}$ satisfies specific regularity properties, ensuring the well-posedness of (1.2.5) in the viscosity sense. To show the convergence of U^ε , the main idea is based on an argument reminiscent of the perturbed test function method, introduced by Evans [Eva89], though applied here in the context of equations over $\mathcal{P}(\mathbb{T}^d)$.

1.3 Model calibration via martingale optimal transport and martingale Schrödinger bridges

Since the introduction of the Black-Scholes model [BS73], a lot of effort has been put into developing sophisticated volatility models that properly capture the market dynamics. To correct for any systematic errors and ensure accurate predictions, it is necessary to properly adjust these models to Vanilla options⁷, such as an S&P500 (SPX) option, which depends on

7. The term Vanilla options refers to financial contracts that give the holder the right to buy or sell an asset at a predetermined price within a specific time frame.

the eponymous stock market index⁸, and VIX options (options on the volatility index). These problems are called calibration problems and are the subject of intense current research; we refer to the survey [GLOW21] and the references therein. To calibrate stochastic volatility models, many authors have employed the tools of optimal transport and Schrödinger bridges, which are related to the transportation of a probability measure to another in an optimal way. The optimization is with respect to a given cost functional and the relative entropy, respectively. Moreover, due to the arbitrage-free nature of many studies, additional martingale constraints are considered. We give an overview of these problems and outline their role in the calibration of stochastic volatility models.

Let $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R})$ be two Borel probability measures in convex order⁹ $\mu_1 \leq_c \mu_2$ and $T_2 > T_1 > 0$. In the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t), \mathbb{P}_0)$, with $\Omega := C([0, T_2]; \mathbb{R}^2)$, we consider the continuous time stochastic volatility model

$$\begin{cases} dX_t = \sigma(X_t, Y_t) dW_t, \\ dY_t = b(X_t, Y_t) dt + \tau_1(X_t, Y_t) dW_t + \tau_2(X_t, Y_t) dW_t^\perp, \end{cases} \quad (1.3.1)$$

where $W, W^\perp$ are independent $\mathbb{P}_0$ -Brownian motions and $\sigma, b, \tau_1, \tau_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$. In the mathematical finance literature, known stochastic volatility models captured by this general form can be the Heston model and the SABR model. In (1.3.1), X_t represents the price of S&P500, while Y_t is the value of its volatility. Calibrating (1.3.1) to SPX options, is equivalent to finding a probability measure $\mathbb{P}$ in $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t))$, such that the laws $\mathcal{L}^\mathbb{P}(X_{T_1}) = \mu_1$ and $\mathcal{L}^\mathbb{P}(X_{T_2}) = \mu_2$ ¹⁰. At the same time, we require X_t to be a $\mathbb{P}$ -martingale and $\mathbb{P}$ to be close to $\mathbb{P}_0$, so that the model remains as close as possible to the initial belief (1.3.1).

8. The Standard and Poor's 500 index, or simply the S&P 500, tracks the performance of 500 of the largest publicly traded companies in the United States.

9. This means that $\int h(x) d\mu_1(x) \leq \int h(x) d\mu_2(x)$ for any convex function $h : \mathbb{R} \rightarrow \mathbb{R}$.

10. In this case T_1, T_2 are the maturities of the options, while the constraints on the law also give the prices of the options with maturities T_1, T_2 : $\mathbb{E}^\mathbb{P}[(X_{T_i} - a)_+] = \int (x - a)_+ d\mu_i(x)$, $i = 1, 2$, for all strike prices $a \in \mathbb{R}$.

In many works, the role of martingale optimal transport and martingale Schrödinger bridges is crucial, especially in quantifying how close $\mathbb{P}, \mathbb{P}_0$ will be. In particular, it is important to establish the probability measure $\mathbb{P}$ which is closest to $\mathbb{P}_0$ in the appropriate sense. These minimization problems often admit duality formulas, which are important in the numerical approximation of the optimal $\mathbb{P}^*$, whenever it exists.

In Chapter 4, we first give a new proof for a duality formula for calibration via martingale optimal transport that can be found in [GLW17] with moment constraints. Let $T_0 \in [0, T_2]$ and $(X_t)_{t \in [T_0, T_2]}$ a continuous time stochastic process with values in $\mathbb{R}$ over the filtered space $(\Omega, \mathcal{F}, \mathbb{F})$. We consider the minimization problem

$$V(T_0, \mu, \mu_1, \mu_2) := \inf_{\mathbb{P} \in \mathcal{P}_m} \left\{ \mathbb{E}^{\mathbb{P}} \left[\int_{T_0}^{T_2} L(\sigma_t \sigma_t^\top) ds \right] \right\}, \quad (1.3.2)$$

where $L : \mathbb{R} \rightarrow \mathbb{R}$ is a given function. The minimization is over the set of probability measures in $(\Omega, \mathcal{F}, \mathbb{F})$

$$\mathcal{P}_m = \left\{ \mathbb{P} \mid X_t \text{ is a } \mathbb{P}\text{-martingale and } \mathcal{L}^{\mathbb{P}}(X_{T_i}) = \mu_i, i = 0, 1, 2 \right\}. \quad (1.3.3)$$

We note that X_t , being a $\mathbb{P}$ -martingale, satisfies $dX_t = \sigma_t dB_t$ with $(B_t)_{t \geq 0}$ being a 1-dimensional Brownian motion over the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ and σ_t an adapted process with values in $\mathbb{R}$.

Theorem 1.3.1. *Suppose that L has at least quadratic growth. Let $\mu_0, \mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R})$ be in convex increasing order ($\mu_0 \leq_c \mu_1 \leq_c \mu_2$), $0 < T_1 < T_2$ and $T_0 \in [0, T_2]$. Then, (1.3.2) is admissible. Furthermore, whenever V is finite, the following duality formula holds*

$$\begin{aligned} V(T_0, \mu_0, \mu_1, \mu_2) &= \sup_{u_1, u_2 \in Lip} \left\{ \int u(T_0, x) d\mu_0(x) - \int u_2(x) d\mu_2(x) \right. \\ &\quad \left. - \mathbf{1}_{[0, T_1]}(T_0) \int u_1(x) d\mu_1(x) \right\}, \end{aligned} \quad (1.3.4)$$

where the supremum is taken over all bounded Lipschitz functions u_1, u_2 , u is a solution (in the appropriate sense), of the Hamilton-Jacobi equation

$$\begin{cases} -\partial_t u + H\left(\frac{D^2 u}{2}\right) = \delta_{T_1}(t)u_1, & (t, x) \in [T_0, T_2] \times \mathbb{R}^d, \\ u(T_2, x) = u_2(x), & x \in \mathbb{R}^d, \end{cases} \quad (1.3.5)$$

and $H(a) = \sup_{b \geq 0} \{-ab - L(b)\}$, for any $a \in \mathbb{R}$, is the Hamiltonian. Moreover, the same formula holds if the supremum in (1.3.4) is taken over u, u_1, u_2 such that u is a classical super-solution of (1.3.5).

We then give a rigorous proof of a relaxation of the duality formula from [Guy22] that refers to calibration via martingale Schrödinger bridges when there is a VIX constraint. In this case, we consider (1.3.1) with initial condition $(X_{t_0}, Y_{t_0}) \in (0, +\infty) \times \mathbb{R}$ at time $t_0 \in [0, T_1]$. The minimization problem is

$$D_{\mathbb{P}_0} := \inf_{\mathbb{P} \in \mathcal{C}_{t_0}(\mu_1, \mu_2), \text{VIX}(G)} H(\mathbb{P}|\mathbb{P}_0), \quad (1.3.6)$$

where H is the relative entropy between $\mathbb{P}$ and $\mathbb{P}_0$ ¹¹ and

$$\mathcal{C}_{t_0}(\mu_1, \mu_2) = \left\{ \mathbb{P} \in \mathcal{P} \left| \mathcal{L}^{\mathbb{P}}(X_{T_i}) = \mu_i, i = 1, 2 \text{ and } X \text{ is an } ((\mathcal{F}_t), \mathbb{P})\text{-martingale} \right. \right\}. \quad (1.3.7)$$

The $\text{VIX}(G)$ is a constraint on the distribution of $V'_{\mathbb{P}} = \mathbb{E}^{\mathbb{P}} \left[G(X_{T_2}) - G(X_{T_1}) \middle| \mathcal{F}_{T_1} \right]$ for some function $G : [0, +\infty) \rightarrow \mathbb{R}$. In particular, we require $\mathcal{L}^{\mathbb{P}}(V'_{\mathbb{P}}) \leq_c \mu_3$ in convex order, where μ_3 is a given probability distribution.

Theorem 1.3.2. *Assume $\sigma, b, \tau_1, \tau_2$ are C^2 and Lipschitz. In addition, we suppose that $|Db| + |D\tau_2| \leq \lambda_0 |\tau_2|$ for some $\lambda_0 > 0$ and that σ is bounded with $\sigma(0, y) = 0$ for every $y \in \mathbb{R}$.*

11. The relative entropy between two probability measures μ, ν is defined as follows:

$$H(\mu|\nu) = \begin{cases} \int \log \frac{d\mu}{d\nu} d\mu, & \text{if } \mu \ll \nu, \\ +\infty, & \text{otherwise.} \end{cases}$$

Suppose (X_t, Y_t) has the dynamics (1.3.1) with initial condition $X_{t_0} > 0$ and $Y_{t_0} > 0$ and that μ_1, μ_2, μ_3 are three probability distributions over $\mathbb{R}$ such that $\mu_1 \leq_c \mu_2$ and $\int x^2 d\mu_2(x) < \infty$. We also assume that (1.3.6) is admissible and G is bounded, convex and Lipschitz over $[0, +\infty)$. Let u be the unique viscosity solution of

$$\begin{cases} -\partial_t u - \mathcal{L}_{x,y}^0 u + \frac{1}{2} \tau_2^2(x, y) (\partial_y u)^2 = 0, & t \in (T_1, T_2], (x, y) \in [0, +\infty) \times \mathbb{R}, \\ u(T_2, x, y) = u_2(x) + \delta G(x), & (x, y) \in [0, +\infty) \times \mathbb{R}, \\ u(t, 0, y) = u_2(0) + \delta G(0), & t \in (T_1, T_2], y \in \mathbb{R}, \end{cases} \quad (1.3.8)$$

for $\delta \in \mathbb{R}$. Let $\Phi(x, y) = \inf_{v \geq 0} \sup_{\delta \in \mathbb{R}} \{u_3(v) - \delta(G(x) + v) + u(T_1^+, x, y; \delta)\}$. Then, Φ is bounded and we have for $t_0 < T_1$

$$\inf_{\mathcal{C}_{t_0}(\mu_1, \mu_2), VIX(G)} H(\mathbb{P}|\mathbb{P}_0) = \sup_{u_1, u_2, u_3} \left\{ - \int u_1 d\mu_1 - \int u_2 d\mu_2 - \int u_3 d\mu_3 + u(t_0, X_{t_0}, Y_{t_0}) \right\}, \quad (1.3.9)$$

where the supremum is taken over all bounded and Lipschitz u_1, u_2 , and u_3 is convex and lower bounded, and where $u(t, x, y)$, $t \in [t_0, T_1]$ is a viscosity solution of

$$\begin{cases} -\partial_t u - \mathcal{L}_{x,y}^0 u + \frac{1}{2} \tau_2^2(x, y) (\partial_y u)^2 = 0, & t \in [t_0, T_1], (x, y) \in [0, +\infty) \times \mathbb{R}, \\ u(T_1, x, y) = u_1(x) + \Phi(x, y), & (x, y) \in [0, +\infty) \times \mathbb{R}, \\ u(t, 0, y) = u_1(0) + \Phi(0, y), & t \in [t_0, T_1], y \in \mathbb{R}. \end{cases} \quad (1.3.10)$$

Here, $\mathcal{L}_{x,y}^0$ is the generator of (1.3.1).

CHAPTER 2

HOMOGENIZATION OF SLOW-FAST CONDITIONAL MCKEAN-VLASOV SDES

2.1 Introduction

Let $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1, \{\mathcal{F}_t^1\})$, $(\Omega^0, \mathcal{F}^0, \mathbb{P}^0, \{\mathcal{F}_t^0\})$ be two filtered probability spaces with filtrations $\{\mathcal{F}_t^1\}$, $\{\mathcal{F}_t^0\}$, B_t, W_t independent standard m -dimensional $\{\mathcal{F}_t^1\}$ -Brownian motions and B_t^0 an m -dimensional $\{\mathcal{F}_t^0\}$ Brownian motion. For any parameter $\lambda \in \mathbb{R}$ small enough, we study the system of slow-fast McKean-Vlasov stochastic differential equations (SDEs for short)

$$\begin{aligned}
dX_t^\epsilon &= \left(\frac{\lambda}{\epsilon} f(X_t^\epsilon, Y_t^\epsilon, \mathcal{L}(X_t^\epsilon | \mathcal{F}_t^0)) + c(X_t^\epsilon, Y_t^\epsilon, \mathcal{L}(X_t^\epsilon | \mathcal{F}_t^0)) \right) dt \\
&\quad + \sigma(X_t^\epsilon, Y_t^\epsilon, \mathcal{L}(X_t^\epsilon | \mathcal{F}_t^0)) dW_t + \sigma_1^0(X_t^\epsilon, \mathcal{L}(X_t^\epsilon | \mathcal{F}_t^0)) dB_t^0, \\
dY_t^\epsilon &= \frac{1}{\epsilon} \left(\frac{1}{\epsilon} f(X_t^\epsilon, Y_t^\epsilon, \mathcal{L}(X_t^\epsilon | \mathcal{F}_t^0)) + g(X_t^\epsilon, Y_t^\epsilon, \mathcal{L}(X_t^\epsilon | \mathcal{F}_t^0)) \right) dt \\
&\quad + \frac{1}{\epsilon} \left(\tau_1(X_t^\epsilon, Y_t^\epsilon, \mathcal{L}(X_t^\epsilon | \mathcal{F}_t^0)) dW_t \right. \\
&\quad \left. + \tau_2(X_t^\epsilon, Y_t^\epsilon, \mathcal{L}(X_t^\epsilon | \mathcal{F}_t^0)) dB_t + \sigma_2^0(X_t^\epsilon, \mathcal{L}(X_t^\epsilon | \mathcal{F}_t^0)) dB_t^0 \right),
\end{aligned} \tag{2.1.1}$$

where $\mathcal{L}(X_t^\epsilon | \mathcal{F}_t^0)$ is the conditional law of X_t^ϵ and where $b, c, f, g : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^d$, $\sigma, \tau_1, \tau_2 : \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^{d \times m}$, $\sigma_1^0, \sigma_2^0 : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^{d \times m}$ $\eta \in L^2(\Omega^1, \mathcal{F}^1, \mathbb{P}^1; \mathbb{R}^d)$ with $X_0 = \eta \sim \mu \in \mathcal{P}_2(\mathbb{R}^d)^1$, $Y_0 = \zeta \in L^p(\Omega^1, \mathcal{F}^1, \mathbb{P}^1; \mathbb{R}^d)$ for all $p > 0$, and (η, ζ) is independent of (W, B) . We note that this also means that $(\eta, \zeta) \sim m \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ such that the y -marginal of m has finite p -moments for every $p > 0$. Here $\mathcal{P}_2(\mathbb{R}^n)$, $n \geq 0$, denotes the space of probability measures over $\mathbb{R}^n$ with finite second moment, equipped with the 2-Wasserstein metric.

The theory of averaging for diffusion process has a long history and broad applications in

1. We use the symbol $\eta \sim \mu$ to denote that the law of η is μ , that is $\mathcal{L}(\eta) = \mu$.

various fields, such as materials sciences, chemistry, biology, and climate dynamics, among others (see e.g [WRD12], [MHR07], [WLVE05], [HKWS11]). Existing results of averaging for SDEs with coefficients that do not depend on the law of the solution are numerous two of which are by [PV01a, PV03] and [RSX19b]. So far in the literature for slow-fast McKean-Vlasov SDEs, when the coefficients depend also on the law of the solution, but there is no common noise, the problem has been studied, with different methods, by [BS23b], [HLS22], [LWX22a],[QW22a] and [RX21a].

The aim of this paper is to describe the behavior of the slow component X_t^ϵ of (2.1.1) as $\epsilon \rightarrow 0$, in presence of the common noise B_t^0 , and establish that it converges to an averaged process, which we will characterize.

Let $a(x, y, \mu) = \frac{1}{2}[\tau_1 \tau_1^\top + \tau_2 \tau_2^\top + \sigma_2^0 \sigma_2^{0\top}](x, y, \mu)$. For fixed $x \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, we consider the operator $L_0^{x,\mu}$ acting on $\phi \in C_b^2(\mathbb{R}^d)$ (that is ϕ has derivatives up to order two which are bounded) with

$$L_0^{x,\mu} \phi = f(x, y, \mu) \cdot D_y \phi(y) + a(x, y, \mu) : D_{yy}^2 \phi(y). \quad (2.1.2)$$

Under our assumptions (see the corresponding section), by Proposition 1 from [PV01a], there exists a $\pi(\cdot; x, \mu)$ which is the unique probability measure solving the distributional equation

$$(L_0^{x,\mu})^* \pi = 0. \quad (2.1.3)$$

Moreover, all moments of π are bounded uniformly in x, μ . Furthermore, $\Phi_k(x, y, \mu) = -\lambda y_k$, $k \in \{1, \dots, d\}$ is a classical solution to the equation

$$L_0^{x,\mu} \Phi_k(x, y, \mu) = -\lambda f_k(x, y, \mu), \quad k \in \{1, 2, \dots, d\}, \quad x, y \in \mathbb{R}^d, \quad \mu \in \mathcal{P}_2(\mathbb{R}^d), \quad (2.1.4)$$

where f_k , $k \in \{1, \dots, d\}$ is the k -component of f . Now, let

$$\begin{aligned}
\gamma(x, y, \mu) &:= c(x, y, \mu) - \lambda g(x, y, \mu) \\
D(x, y, \mu) &:= D_1(x, y, \mu) + D_1^\top(x, y, \mu) + \frac{1}{2} \sigma(x, y, \mu) \sigma^\top(x, y, \mu) \\
&\quad - \frac{\lambda^2}{2} \sigma_2^0 \sigma_2^{0\top}(x, \mu) \\
D_1(x, y, \mu) &:= -\frac{\lambda}{2} [\lambda f(x, y, \mu) \otimes y + \tau_1(x, y, \mu) \sigma^\top(x, y, \mu)] \\
\Sigma(x, \mu) &:= \sigma_1^0(x, \mu) - \lambda \sigma_2^0(x, \mu)
\end{aligned} \tag{2.1.5}$$

and

$$\bar{\gamma}(x, \mu) := \int_{\mathbb{R}^d} \gamma(x, y, \mu) \pi(dy; x, \mu), \quad \bar{D}(x, \mu) := \int_{\mathbb{R}^d} D(x, y, \mu) \pi(dy; x, \mu). \tag{2.1.6}$$

In this paper, we will study the convergence as well as the rate of convergence, of $\mathcal{L}(X_t^{\epsilon, m} | \mathcal{F}_t^0)$ to $\mathcal{L}(X_t^\mu | \mathcal{F}_t^0)$ on the space $\mathcal{P}_2(\mathbb{R}^d)$, where X_t^μ satisfies the averaged conditional McKean-Vlasov SDE:

$$\begin{aligned}
X_t^\mu &= \eta^2 + \int_0^t \bar{\gamma}(X_s^\mu, \mathcal{L}(X_s^\mu | \mathcal{F}_s^0)) ds + \int_0^t \sqrt{2\bar{D}}^{1/2}(X_s^\mu, \mathcal{L}(X_s^\mu | \mathcal{F}_s^0)) dW_s^2 \\
&\quad + \int_0^t \Sigma(X_s^\mu, \mathcal{L}(X_s^\mu | \mathcal{F}_s^0)) dB_s^0.
\end{aligned} \tag{2.1.7}$$

The equation is posed on a possibly different filtered probability space supporting a d -dimensional Brownian motion W^2 , $X_0^{\epsilon, \mu} = \eta^2$ is random variable on this new probability space independent from W^2 with $\eta^2 \sim \mu \in \mathcal{P}_2(\mathbb{R}^d)$. Finally, $\bar{D}^{1/2}(x, \mu)$ is the unique positive semi-definite matrix such that $\bar{D}^{1/2}(x, \mu) \bar{D}^{1/2}(x, \mu) = \bar{D}(x, \mu)$. Note that it is not clear that $\bar{D}$ is symmetric and positive semi-definite, therefore we will include it in our assumptions.

2.1.1 Notation

- For any two matrices A, B we denote $A : B = \text{tr}(AB^\top)$.
- For $i = 1, 2$, $\pi^i : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ is the projection $\pi^i(x_1, x_2) = x_i$.
- Let X, Y and Z ($= \mathbb{R}$ or $\mathbb{R}^d$) be metric spaces. For any function $h : X \times Y \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow Z$ we set $\tilde{h}(x, y, m) = h(x, y, \pi_*^1 m)$ for any $m \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$.
- If $X, Y, Z = \mathbb{R}^d$, we also write $\bar{h}(x, \mu) = \int h(x, y, \mu) \pi_{x, \mu}(dy)$, where $\pi_{x, \mu}$ is the probability measure solving (2.1.3).
- We denote by $\mathcal{A}$ the set of all $m \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ such that the pushforward measure $\pi_*^2 m$ has finite p -moments for every $p > 0$. We note that $\mathcal{A}$ is a convex subset of $\mathcal{P}_2$.
- For $s \in [t, T]$, we will use the symbol $(X_s^\epsilon, Y_s^\epsilon)^{m, t} = (X_s^{\epsilon, m, t}, Y_s^{\epsilon, m, t})$ for the solution of (2.1.1) with initial joint distribution (at $s = t$) equal to m . Similarly, we define $X_s^{\mu, t}$ for the solution of (2.1.7).
- We define the product structure $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$, where $\Omega = \Omega^1 \times \Omega^0$, $(\mathcal{F}, \mathbb{P})$ is the completion of $(\mathcal{F}^1 \otimes \mathcal{F}^0, \mathbb{P}^1 \otimes \mathbb{P}^0)$ and $\{\mathcal{F}_t\}$ is the complete right continuous augmentation of $\{\mathcal{F}_t^1 \otimes \mathcal{F}_t^0\}$.
- For $i = 1, 2$, we will denote by $\mathbb{E}^i$ the integration with respect to $\mathbb{P}^i$ and by $\mathbb{E}$ the integration with respect to $\mathbb{P}$.
- For the operator $L_0^{x, \mu}$ (or later $L_{0,2}^{x, x', \mu}$) we will sometimes abuse the notation and write $L_0^{x, \mu} \phi$ even if ϕ is vector or matrix valued meaning that we apply $L_0^{x, \mu}$ to every coordinate or entry of ϕ .
- If $U : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$, its Wasserstein derivative is denoted by $\frac{\delta U}{\delta m}(m, x)$ for $m \in \mathcal{P}_2(\mathbb{R}^d)$, $x \in \mathbb{R}^d$. We refer to 2.2.
- For a function $u : [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$, we use the notation $\dot{u}(t, x)$ for the time derivative $\partial_t u(t, x)$, while we use $D_x u(t, x)$ for the space derivative.

2.1.2 The main result

The aforementioned convergence result involves test functions G which are at least 4 times continuously differentiable. In particular, we consider the function

$$u^\epsilon(t, m) = \mathbb{E}^0[G(\mathcal{L}(X_T^{\epsilon, t, m} | \mathcal{F}_T^0))] \quad (2.1.8)$$

and we show that $u^\epsilon(t, m) \xrightarrow{\epsilon \rightarrow 0} u(t, \pi_*^1 m)$, where u satisfies, in the classical sense, the following equation which is posed over the Wasserstein space $\mathcal{P}_2(\mathbb{R}^d)$:

$$\begin{cases} \dot{u}(t, \mu) + \mathcal{L}u(t, \mu) = 0, & (t, \mu) \in [0, T) \times \mathcal{P}_2(\mathbb{R}^d) \\ u(T, \mu) = G(\mu), & \mu \in \mathcal{P}_2(\mathbb{R}^d), \end{cases} \quad (2.1.9)$$

with

$$\begin{aligned} \mathcal{L}u(t, \mu) = & \int_{\mathbb{R}^d} \left(\bar{\gamma}(x, \mu) \cdot D_\mu u(t, \mu, x) + (\bar{D}(x, \mu) + \frac{1}{2} \Sigma \Sigma^\top(x, \mu)) : D_x D_\mu u(t, \mu, x) \right) \mu(dx) \\ & + \frac{1}{2} \int_{\mathbb{R}^{2d}} \Sigma(x, \mu) \Sigma^\top(x', \mu) : D_{\mu\mu}^2 u(t, \mu, x, x') \mu(dx) \mu(dx'). \end{aligned} \quad (2.1.10)$$

In Theorem 2.7.1, we establish that (2.1.9) has a unique classical solution which has the formula $u(t, \mu) = \mathbb{E}^0[G(\mathcal{L}(X_T^{t, \mu} | \mathcal{F}_T^0))]$, where $X_T^{t, \mu}$ is the averaged process (2.1.7). We note that this is an infinite dimensional analogue of the Feynmann-Kac formula. Such formulas have also been established in the case of McKean-Vlasov SDEs without common noise in [BDL17, CST22]. To the author's knowledge, this is the first time that the PDE associated with conditional McKean-Vlasov SDEs is studied in this context.

With these in mind, the exact statement of the homogenization result is the following.

Theorem 2.1.1. *Assume Hypotheses (H1)-(H6) and let $T \in [0, +\infty)$, $m \in \mathcal{A}$ and $t \in [0, T]$. We consider $\dot{\zeta}$ the set of all multi-indices (n, β) with $|(n, \beta)| \leq 4$. Then, there exists a*

constant $C(T)$, independent of $\mu = \pi_*^1 m$, such that for any $G \in \mathcal{M}_{b,L}^\dot{\xi}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$

$$\sup_{t \in [0, T]} |u^\epsilon(t, m) - u(t, \mu)| \leq \epsilon C(T) \|G\|_{\mathcal{M}_{b,L}^\dot{\xi}},$$

where u is the unique classical solution of (2.1.9) and u^ϵ is as in (2.1.8). Equivalently,

$$\sup_{t \in [0, T]} \left| \mathbb{E}^0 \left[G(\mathcal{L}(X_T^{\epsilon, m, t} | \mathcal{F}_T^0)) - G(\mathcal{L}(X_T^{\mu, t} | \mathcal{F}_T^0)) \right] \right| \leq \epsilon C(T) \|G\|_{\mathcal{M}_{b,L}^\dot{\xi}}. \quad (2.1.11)$$

The idea of the proof is to show that the solution of the McKean-Vlasov control equation² associated with (2.1.1) converges to the solution of the McKean-Vlasov control equation (2.1.9) associated with the averaged process (2.1.7). Indeed, as discussed in [PW17] the McKean-Vlasov control equation associated with the system (2.1.1) is

$$\begin{cases} \dot{u}^\epsilon(t, m) + \mathcal{L}^\epsilon u^\epsilon(t, m) = 0, & \text{in } [0, T) \times \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d), \\ u^\epsilon(T, m) = G(\pi_*^1 m), & \text{in } \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d), \end{cases} \quad (2.1.12)$$

where

$$\mathcal{L}^\epsilon u(t, m) = \mathcal{L}_2 u + \frac{1}{\epsilon} \mathcal{L}_1 u + \frac{1}{\epsilon^2} \mathcal{L}_0 u, \quad \text{with}$$

$$\mathcal{L}_2 u = \int_{\mathbb{R}^{2d}} \int_{\mathbb{R}^{2d}} \left(\frac{\widetilde{\sigma_1^0}(x, m) \widetilde{\sigma_1^{0\top}}(x', m)}{2} : D_{xx'}^2 \frac{\delta^2 u}{\delta m^2} \right) m(dx, dy) m(dx', dy')$$

2. McKean-Vlasov control equations are studied in the context of Mean Field Control theory. In our case u^ϵ is the value function of a Mean Field Control problem in a trivial sense $u^\epsilon(t, m) = \inf_\alpha \mathbb{E}^0[G(\mathcal{L}(X_T^{\epsilon, t, m} | \mathcal{F}_T^0))]$, therefore the McKean-Vlasov control PDE is linear (see (2.1.12)). We keep this name to underline the connection with Mean Field Control theory.

$$\begin{aligned}
& + \int_{\mathbb{R}^{2d}} \left(\widetilde{c}(x, y, m) \cdot D_x \frac{\delta u}{\delta m} + \frac{\widetilde{\sigma \sigma^\top}(x, y, m)}{2} : D_{xx}^2 \frac{\delta u}{\delta m} \right. \\
& \quad \left. + \frac{\widetilde{\sigma_1^0 \sigma_1^{0\top}}(x, m)}{2} : D_{xx}^2 \frac{\delta u}{\delta m} \right) m(dx, dy), \tag{2.1.13}
\end{aligned}$$

$$\begin{aligned}
\mathcal{L}_1 u &= \frac{1}{2} \int_{\mathbb{R}^{2d}} \int_{\mathbb{R}^{2d}} \left(\widetilde{\sigma_2^0}(x, m) \widetilde{\sigma_1^{0\top}}(x', m) : D_{x'y}^2 \frac{\delta^2 u}{\delta m^2} \right) m(dx, dy) m(dx', dy') \\
& + \frac{1}{2} \int_{\mathbb{R}^{2d}} \int_{\mathbb{R}^{2d}} \left(\widetilde{\sigma_1^0}(x, m) \widetilde{\sigma_2^{0\top}}(x', m) : D_{xy'}^2 \frac{\delta^2 u}{\delta m^2} \right) m(dx, dy) m(dx', dy') \\
& + \int_{\mathbb{R}^{2d}} \left(\widetilde{f}(x, y, m) \cdot D_x \frac{\delta u}{\delta m} + \widetilde{g}(x, y, m) \cdot D_y \frac{\delta u}{\delta m} \right. \\
& \quad \left. + \widetilde{\tau_1 \sigma^\top}(x, y, m) : D_{xy}^2 \frac{\delta u}{\delta m} + \widetilde{\sigma_1^0 \sigma_2^{0\top}}(x, m) : D_{yx}^2 \frac{\delta u}{\delta m} \right) m(dx, dy), \tag{2.1.14}
\end{aligned}$$

$$\begin{aligned}
\mathcal{L}_0 u &= \int_{\mathbb{R}^{2d}} \int_{\mathbb{R}^{2d}} \left(\frac{\widetilde{\sigma_2^0}(x, m) \widetilde{\sigma_2^{0\top}}(x', m)}{2} : D_{yy'}^2 \frac{\delta^2 u}{\delta m^2} \right) m(dx, dy) m(dx', dy') \\
& + \int_{\mathbb{R}^{2d}} \left(\widetilde{f}(x, y, m) \cdot D_y \frac{\delta u}{\delta m} + \frac{(\widetilde{\tau_1 \tau_1^\top} + \widetilde{\tau_2 \tau_2^\top})(x, y, m)}{2} : D_{yy}^2 \frac{\delta u}{\delta m} \right. \\
& \quad \left. + \frac{\widetilde{\sigma_2^0 \sigma_2^{0\top}}(x, m)}{2} : D_{yy}^2 \frac{\delta u}{\delta m} \right) m(dx, dy) \tag{2.1.15}
\end{aligned}$$

and the terminal condition G depends only on the x -marginal of m .

We know that u^ϵ from (2.1.8) is a solution of (2.1.12). Now, let μ be the x -marginal of m . We are expecting that, heuristically, under our assumptions on the coefficients, $u^\epsilon(t, m)$ will have an expansion of the form $u(t, \mu) + \epsilon u_1(t, m) + \epsilon^2 u_2(t, m) = u(t, \mu) + \mathcal{O}(\epsilon)$, where u_1, u_2 will be some corrector functions. If $\epsilon \rightarrow 0$, this will converge to $u(t, \mu)$, which gives us the desired convergence and the rate of convergence. The proof relies heavily on the regularity properties of the solutions of the McKean-Vlasov control equation (2.1.12) and the Poisson equation associated with the operator $\mathcal{L}_0$ from (2.1.15).

2.1.3 Summary of the chapter

In section 2, we discuss the Wasserstein space and Wasserstein differentiation. In section 3, we state our assumptions on the coefficients as well as some preliminary properties related to (2.1.1). Sections 4 and 5 contain some results on the McKean-Vlasov control equation and the Poisson problem, which will be useful in the proof. In section 6, we first construct the correctors u_1, u_2 and we use the results from sections 4 and 5 to establish bounds on them and their derivatives. Finally, we use the properties of u_1, u_2 to finish the proof of Theorem 2.1.1. We use the last two sections to present the technical proofs and some remarks on the main result of this paper.

2.2 Wasserstein space and Wasserstein differentiation

This section is devoted to the Wasserstein space and the notion of differentiation defined on it. We will focus on the Wasserstein space $\mathcal{P}_2$, even though the results can be extended to $\mathcal{P}_p$, $p > 0$. Let $m_1, m_2 \in \mathcal{P}_2(\mathbb{R}^n)$ the metric we will be using is given by

$$\mathbf{d}_2(m_1, m_2) = \inf_{\substack{\gamma \in \mathcal{P}_2(\mathbb{R}^n \times \mathbb{R}^n) \\ \pi_*^1 \gamma = m_1, \pi_*^2 \gamma = m_2}} \left(\int_{\mathbb{R}^{2n}} |x - y|^2 \gamma(dx, dy) \right)^{1/2}.$$

Definition 2.2.1. Let $U : \mathcal{P}_2(\mathbb{R}^n) \rightarrow \mathbb{R}$, $n \geq 1$.

(i) We say that U is a C^1 function if there exists a continuous map $K : \mathcal{P}_2(\mathbb{R}^n) \times \mathbb{R}^n \rightarrow \mathbb{R}$ such that for all $m_1, m_2 \in \mathcal{P}_2(\mathbb{R}^n)$ we have $\int_{\mathbb{R}^n} K(m_1, x) m_2(dx) < \infty$ and

$$\lim_{t \rightarrow 0} \frac{U((1-t)m_1 + tm_2) - U(m_1)}{t} = \int_{\mathbb{R}^n} K(m_1, x)(m_2 - m_1)(dx).$$

We will use the symbol $\frac{\delta U}{\delta m}(m, x)$ for the map $K(m, x)$, which is called the Wasserstein flat derivative.

(ii) If U is of class C^1 and $\frac{\delta U}{\delta m}$ is C^1 with respect to the space variable, we define the

intrinsic derivative (Wasserstein derivative) $D_m U : \mathcal{P}_2(\mathbb{R}^n) \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ as $D_m U(m, x) := D_x \frac{\delta U}{\delta m}(m, x)$.

Remark 2.2.2. (i) Note that in the first part of the above definition, the Wasserstein flat derivative is defined up to a constant.

(ii) We can adopt the same definition for a function U defined on a convex subset of $\mathcal{P}_2(\mathbb{R}^n)$ with values in $\mathbb{R}$. In particular, U may be defined on the set $\mathcal{A} \subset \mathcal{P}_2(\mathbb{R}^{2d})$ of all measures whose second marginal $\pi_2^* m$ (see below) has finite moments.

We continue with some properties of the functions defined over $\mathcal{P}_2(\mathbb{R}^n)$, $n \geq 1$, in relation to composition with the projections $\pi^1, \pi^2 : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ with $\pi^1(x_1, x_2) = x_1$ and $\pi^2(x_1, x_2) = x_2$. It is known that if $m \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$, then $\pi_*^1 m$ and $\pi_*^2 m$ are the x and y marginals, respectively. In particular, $\pi_*^i : \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d) \rightarrow \mathcal{P}_2(\mathbb{R}^d)$.

Proposition 2.2.3. *Let $m_1, m_2 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$. Then, the following inequality is true*

$$\max \{ \mathbf{d}_2(\pi_*^1 m_1, \pi_*^1 m_2), \mathbf{d}_2(\pi_*^2 m_1, \pi_*^2 m_2) \} \leq \mathbf{d}_2(m_1, m_2).$$

Proof. For any $\gamma \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d)$ with marginals m_1 and m_2 we have

$$\begin{aligned} \int |(x_1, y_1) - (x_2, y_2)|^2 \gamma(dx_1, dx_2, dy_1, dy_2) &\geq \int |x_1 - x_2|^2 \gamma(dx_1, dx_2, dy_1, dy_2) \\ &= \int |x_1 - x_2|^2 \gamma_1(dx_1, dx_2) \geq \mathbf{d}_2(\pi_*^1 m_1, \pi_*^1 m_2)^2, \end{aligned}$$

where γ_1 has marginals $\pi_*^1 m_1$ and $\pi_*^1 m_2$. Therefore, $\mathbf{d}_2(m_1, m_2) \geq \mathbf{d}_2(\pi_*^1 m_1, \pi_*^1 m_2)$. Similarly, we can show that $\mathbf{d}_2(m_1, m_2) \geq \mathbf{d}_2(\pi_*^2 m_1, \pi_*^2 m_2)$.

□

The following propositions describe how the regularity of functions defined on $\mathcal{P}_2(\mathbb{R}^d)$ change when we compose with the projections π_*^i , $i = 1, 2$.

Proposition 2.2.4. *Let $U : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ be a Lipschitz function with Lipschitz constant equal to L . Then, for $i = 1, 2$ the function $\tilde{U}_i : \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d) \rightarrow \mathbb{R}$ with $\tilde{U}_i(m) = U(\pi_*^i m)$ is also Lipschitz with the same Lipschitz constant.*

Proof. Indeed, let $m_1, m_2 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$. Then, for $i = 1, 2$ we have

$$|\tilde{U}(m_1) - \tilde{U}(m_2)| = |U(\pi_*^i m_1) - U(\pi_*^i m_2)| \leq L \mathbf{d}_2(\pi_*^i m_1, \pi_*^i m_2) \leq L \mathbf{d}_2(m_1, m_2).$$

□

Proposition 2.2.5. *Let $U : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ be a C^1 function. Then, for $i = 1, 2$ the function $\tilde{U}_i : \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d) \rightarrow \mathbb{R}$ with $\tilde{U}_i(m) = U(\pi_*^i m)$ is also C^1 with*

$$\frac{\delta \tilde{U}_i}{\delta m}(m, x_1, x_2) = \frac{\delta U}{\delta m}(\pi_*^i m, \pi^i(x_1, x_2)), \text{ for every } m \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d), x_1, x_2 \in \mathbb{R}^d. \quad (2.2.1)$$

Proof. We will show the result for $i = 1$ as the other case can be shown with an identical argument. For simplicity in the calculations below we drop the index i , so $\tilde{U}_i = \tilde{U}$ and $\pi^1 = \pi$. Let $m_1, m_2 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ and $\pi_* m_1 = \mu_1$, $\pi_* m = \mu^2$. We have by the properties of the pushforward

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{\tilde{U}((1-t)m_1 + tm_2) - \tilde{U}(m_1)}{t} &= \lim_{t \rightarrow 0} \frac{U((1-t)\mu_1 + t\mu_2) - U(\mu_2)}{t} \\ &= \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(\mu_1, x_1)(\mu_2 - \mu_1)(dx_1) \\ &= \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(\pi_* m_1, x_1)(\pi_* m_2 - \pi_* m_1)(dx_1, dx_2) \\ &= \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(\pi_* m_1, x_1)(\pi_* m_2 - \pi_* m_1)(dx_1) \\ &= \int \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{\delta U}{\delta m}(\pi_* m_1, \pi(x_1, x_2))(m_2 - m_1)(dx_1, dx_2). \end{aligned}$$

Equation (2.2.1) follows. The regularity of $\tilde{U}$ is a direct implication of (2.2.1) and the fact that U is C^1 . □

Finally, Proposition 2.2.5 can give us a characterization of a function whose Wasserstein flat derivative is independent of a variable (in analogy to (2.2.1)). Namely,

Proposition 2.2.6. *Let $U : \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d) \rightarrow \mathbb{R}$ be a C^1 function such that $\frac{\delta U}{\delta m}(m, x, y)$ is independent of y , then there exists a C^1 function $\bar{U} : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ such that*

$$U(m) = \bar{U}(\pi_*^1 m), \quad \text{for all } m \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d).$$

Proof. Let $m_0, m_1 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ and set $m_s = (1-s)m_0 + sm_1$, where $s \in [0, 1]$. We also set $\mu_0 = \pi_*^1 m_0$ and $\mu_1 = \pi_*^1 m_1$. Since the flat derivative is independent of y , we observe

$$\begin{aligned} U(m_0) - U(m_1) &= \int_0^1 \int \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{\delta U}{\delta m}(m_s, x, y)(m_1 - m_0)(dx, dy) ds \\ &= \int_0^1 \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(m_s, x)(\mu_1 - \mu_0)(dx) ds, \end{aligned}$$

therefore if $\mu_0 = \mu_1$ we have $U(m_0) = U(m_1)$. This implies that there exists a function $\bar{U}$ defined on $\mathcal{P}_2(\mathbb{R}^d)$ such that $U(m) = \bar{U}(\pi_*^1 m)$. It suffices to show that $\bar{U}$ is C^1 . For any $m_0, m_1 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ we have

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{U((1-t)m_0 + tm_1) - U(m_0)}{t} &= \int \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{\delta U}{\delta m}(m_0, x, y)(m_1 - m_0)(dx, dy) \\ &= \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(m_0, x, y)m_1(dx, dy) \\ &= \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(m_0, x)\mu_1(dx). \end{aligned}$$

Note that if m'_0 is another measure such that $\pi_*^1 m'_0 = \mu_0$, then from this calculation we deduce that

$$\int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(m_0, x)\mu_1(dx) = \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(m'_0, x)\mu_1(dx), \quad \text{for any } \mu_1 \in \mathcal{P}_2(\mathbb{R}^d).$$

Consequently, $\frac{\delta U}{\delta m}(m_0, x)$ is independent of the choice of m_0 such that $\pi_*^1 m_0 = \mu_0$. Thus, for $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ and any $m_0, m_1 \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ such that $\pi_0^1 m_i = \mu_i$, $i = 0, 1$, we have

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{\bar{U}((1-t)\mu_0 + t\mu_1) - \bar{U}(\mu_0)}{t} &= \lim_{t \rightarrow 0} \frac{U((1-t)m_0 + tm_1) - U(m_0)}{t} \\ &= \int \int_{\mathbb{R}^d \times \mathbb{R}^d} \frac{\delta U}{\delta m}(m_0, x, y)(m_1 - m_0)(dx, dy) \\ &= \int_{\mathbb{R}^d} \frac{\delta U}{\delta m}(m_0, x)(\mu_1 - \mu_0)(dx) \end{aligned}$$

and hence $\frac{\delta \bar{U}}{\delta \mu}(\mu, x)$ exists and is equal to $\frac{\delta U}{\delta m}(m, x)$, where $\mu = \pi_*^1 m$. $\square$

2.3 Assumptions and Preliminary results

The probability space $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$ is assumed to be rich enough such that it supports all probability laws on $\mathbb{R}^d$, i.e for any probability law μ on $\mathbb{R}^d$ there exists $\eta : \Omega^1 \rightarrow \mathbb{R}^d$ such that $\mathcal{L}(\eta) = \mu$.

Before we state further assumptions, we introduce the multi-index notation proposed in [CM18]. For the derivatives with respect to the measure argument we use the following definition.

Definition 2.3.1. (L -derivatives) (i) A function $U : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ is said to be L -differentiable if its lifting $U' : L^2(\Omega^1, \mathcal{F}^1, \mathbb{P}^1; \mathbb{R}^d) \rightarrow \mathbb{R}$ with $U'(\eta) := U(\mathcal{L}(\eta))$ admits a continuous Fréchet derivative.

Remark 2.3.2. (i) If U is L -differentiable, then it can be shown that there exists a measurable function $\partial_\mu U(\mu, \cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that the Fréchet derivative of U' at $\eta \in L^2(\Omega^1, \mathcal{F}^1, \mathbb{P}^1; \mathbb{R}^d)$ is equal to $\partial_\mu U(\mu, \eta)$ for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $\eta \in L^2(\Omega^1, \mathcal{F}^1, \mathbb{P}^1; \mathbb{R}^d)$ such that $\mathcal{L}(\eta) = \mu$. We say that $\partial_\mu U$ is the L -derivative of U .

(ii) The definition for higher order L -derivatives is analogous.

(iii) By Propositions 5.48 and 5.51 of [CD⁺18a], if further regularity is assumed, then the L -derivative coincides with the Wasserstein derivative (we refer to 2.2); that is $\partial_\mu U(\mu, x) = D_\mu U(\mu, x) := D_x \frac{\delta U}{\delta \mu}(\mu, x)$. Since the functions with which we will be working satisfy or can be proved that they satisfy the regularity assumptions of these propositions, we will be using both notations interchangeably, without explicitly distinguishing between them.

Definition 2.3.3. (Multi-index notation) Let n, l, k be natural numbers and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_n)$ be an n -dimensional vector of natural numbers. Then, we call any ordered tuple $(n, l, \boldsymbol{\beta})$ or $(n, \boldsymbol{\beta})$ a multi-index. For a function $h : \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^k$, the $(n, l, \boldsymbol{\beta})$ -derivative of h is defined as

$$D^{(n, l, \boldsymbol{\beta})} h(x, m, z_1, \dots, z_n) := D_{z_n}^{\beta_n} \dots D_{z_1}^{\beta_1} D_x^l D_m^n h(x, m, z_1, \dots, z_n),$$

where D_m is the intrinsic Wasserstein derivative. Similarly, for a function $h : \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^k$ we define

$$D^{(n, \boldsymbol{\beta})} h(m, z_1, \dots, z_n) = D_{z_n}^{\beta_n} \dots D_{z_1}^{\beta_1} D_m^n h(m, z_1, \dots, z_n).$$

In the spaces we will define below, we will be using the symbol $\boldsymbol{\zeta}$ for a collection of multi-indices. We will say that $\boldsymbol{\zeta}$ is a **complete** collection of multi-indices if for every $(n, l, \boldsymbol{\beta}) \in \boldsymbol{\zeta}$ we have that $\boldsymbol{\zeta}$ contains all lower-order multi-indices of the same type. For example, if $(2, 0, (1, 1)) \in \boldsymbol{\zeta}$, then $(2, 0, (1, 0))$, $(2, 0, (0, 1))$, $(1, 0, 1)$, $(1, 0, 0)$ and $(0, 0, 0)$ are also in $\boldsymbol{\zeta}$.

Definition 2.3.4. Let $\boldsymbol{\zeta}$ be a collection of multi-indices of the form $(n, l, \boldsymbol{\beta})$ and $k, j, d \in \mathbb{N}$ (k may also be in $\mathbb{N} \times \mathbb{N}$). We define $\mathcal{M}_b^{\boldsymbol{\zeta}}(\mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ to be the class of functions $h : \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^k$ such that the $(n, l, \boldsymbol{\beta})$ -derivative of h exists and satisfies

$$\|h\|_{\mathcal{M}_b^{\boldsymbol{\zeta}}} := \sup_{(n, l, \boldsymbol{\beta}) \in \boldsymbol{\zeta}} \sup_{x \in \mathbb{R}^j, z_1, \dots, z_n \in \mathbb{R}^d, m \in \mathcal{P}_2(\mathbb{R}^d)} |D^{(n, l, \boldsymbol{\beta})} h(x, m, z_1, \dots, z_n)| \leq C.$$

We also write that $h \in \mathcal{M}_{b,L}^\zeta$, if for all $(n, l, \beta) \in \zeta$, $x, x' \in \mathbb{R}^j$, $z_1, \dots, z_n, z'_1, \dots, z'_n \in \mathbb{R}^d$ and $m, m' \in \mathcal{P}_2(\mathbb{R}^d)$

$$|D^{(n,l,\beta)}h(x, m, z_1, \dots, z_n) - D^{(n,l,\beta)}h(x', m', z'_1, \dots, z'_n)| \leq C_L \left(|x - x'| + \sum_{i=1}^n |z_i - z'_i| + \mathbf{d}_2(m, m') \right).$$

We can define the spaces $\mathcal{M}_b^\zeta(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ and $\mathcal{M}_{b,L}^\zeta(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ analogously, where ζ will now be a collection of multi-indices of the form (n, β) , as explained in the previous definition.

Furthermore, we define the class $\mathcal{M}_b^\zeta(\mathbb{R}^j \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ which consists of functions $h : \mathbb{R}^j \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^k$ such that $h(\cdot, y, \cdot) \in \mathcal{M}_b^\zeta$ for all $y \in \mathbb{R}^j$ and for each multi-index $(n, l, \beta) \in \zeta$ there exists $q \in \mathbb{N}$ and $C > 0$ such that

$$\sup_{x \in \mathbb{R}^j, z_1, \dots, z_n \in \mathbb{R}^d, m \in \mathcal{P}_2(\mathbb{R}^d)} |D^{(n,l,\beta)}h(x, y, m, z_1, \dots, z_n)| \leq C(1 + |y|^q)$$

and there exists $\theta \in (0, 1]$ (independent of the multi-index) such that

$$|D^{(n,l,\beta)}h(x, y, m, z_1, \dots, z_n) - D^{(n,l,\beta)}h(x', y', m', z'_1, \dots, z'_n)| \leq C \left(1 \wedge |y - y'|^\theta + |x - x'| + \sum_{i=1}^n |z_i - z'_i| + \mathbf{d}_2(m, m') \right) \left(1 + |y|^q + |y'|^q \right).$$

Moreover, we introduce $\mathcal{M}_b^\zeta([0, T] \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ which is the class of functions $h : [0, T] \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}^k$ such that $h(\cdot, x, m)$ is continuously differentiable on $(0, T)$ for all $x \in \mathbb{R}^j, m \in \mathcal{P}_2(\mathbb{R}^d)$ with time derivative denoted by $\dot{h}(t, x, m)$, $h(t, \cdot, \cdot) \in \mathcal{M}_b^\zeta$ for all $t \in [0, T]$ with $\|h(t, \cdot, \cdot)\|_{\mathcal{M}_b^\zeta} \leq C$ uniformly in t and all the derivatives of h and $\dot{h}$ are jointly continuous with respect to the time, space and measure variables. The norm in this space is

defined as:

$$||h||_{\mathcal{M}_b^\zeta([0,T] \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)} := \sup_{t \in [0,T]} ||h(t, \cdot)||_{\mathcal{M}_b^\zeta} + \sup_{t \in [0,T], x \in \mathbb{R}^j, m \in \mathcal{P}_2(\mathbb{R}^d)} |\dot{h}(t, x, m)|.$$

Finally, we can define the spaces $\mathcal{M}_{b,L}^\zeta([0, T] \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$, $\mathcal{M}_b^\zeta([0, T] \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$, $\mathcal{M}_b^\zeta([0, T] \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ and $\mathcal{M}_p^\zeta([0, T] \times \mathbb{R}^j \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ similarly.

Remark 2.3.5. (i) Notice that for the class $\mathcal{M}_p^\zeta(\mathbb{R}^j \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$ and $\mathcal{M}_p^\zeta([0, T] \times \mathbb{R}^j \times \mathbb{R}^j \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^k)$, we may choose q, C to be independent of the multi-index, because the set of multi-indices is finite. So for a function h in these classes we will write $q = p_h$ and $C = C_h$. For both classes, we thus have $||h||_{\mathcal{M}_p^\zeta} \leq C_h(1 + |y|^{p_h})$.
(ii) In the above definitions we may have $\mathcal{A}$ or any convex subset of $\mathcal{P}_2(\mathbb{R}^d)$ instead of the whole $\mathcal{P}_2(\mathbb{R}^d)$.

2.3.1 Assumptions

Let $\widehat{\zeta}$ be a set of multi-indices (n, l, β) such that $|(n, l, \beta)| = n + l + \beta_1 + \dots + \beta_n \leq 4$. The following hypotheses hold:

(H1) There exist $\alpha, \beta > 0$ such that

$$0 < \alpha|z|^2 \leq z^\top \frac{\tau_1 \tau_1^\top + \tau_2 \tau_2^\top}{2} (x, y, \mu) z \leq \beta|z|^2 < \infty, \text{ for all } x, y, z \in \mathbb{R}^d, z \neq 0, \mu \in \mathcal{P}_2(\mathbb{R}^d)$$

and τ_1, τ_2 are bounded, have two uniformly bounded derivatives in y and both these derivatives are Hölder continuous in y uniformly in (x, μ) .

(H2) There exist constant $C_1, C_2 > 0$ independent of x, y, μ such that

$$f(x, y, \mu) \cdot y \leq -C_1|y|^2 + C_2, \text{ for all } x, y \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d),$$

f grows at most linearly in $|y|$, f has two uniformly bounded derivatives in y and both these derivatives are Hölder continuous in y uniformly in (x, μ) .

(H3) σ_1^0, σ_2^0 are in $\mathcal{M}_b^{\hat{\zeta}}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times m})$.

(H4) The functions $\tau_2, \tau_1, \sigma, c, f, g, \sigma_1^0, \sigma_2^0$ are globally Lipschitz continuous in (x, y, μ) .

In addition g is uniformly bounded.

(H5) $f, g, c \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$, $\sigma, \tau_1, \tau_2 \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times m})$ and $a \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$.

(H6) There exists $\delta > 0$ such that $\overline{D}$ from equation (2.1.6) is positive for every $\lambda \in (-\delta, \delta)$. That is there exists an $\alpha' > 0$ such that $z^\top \overline{D}(x, \mu) z \geq \alpha' |z|^2 > 0$ for every $x, z \in \mathbb{R}^d$, $z \neq 0$, $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ and $\lambda \in (-\delta, \delta)$.

Remark 2.3.6. (i) We note that Assumption (H4) is sufficient for the existence and the uniqueness of a solution of (2.1.1) (see e.g [CD⁺18c] Proposition 2.8).

(ii) Assumptions (H1)-(H3) are useful in the solution of the Poisson equation. Specifically, we use them in Proposition 2.4.1. Assumption (H2) is also used in the proof of Proposition 2.3.7.

(iii) The amount of regularity we have on the coefficients (assumptions (H3), (H4), (H5)) is needed to establish the analogous regularity of the averaged coefficients $(\overline{\gamma}, \overline{D}^{1/2}, \Sigma)$ and, in consequence, of the solution u of the averaged McKean-Vlasov control equation (2.1.9).

(iv) The last assumption (H6) is stated because it is not clear whether $\overline{D}$ is a positive matrix. Due to (H3), (H6) can also follow (for some $\delta > 0$ small) by only assuming that $\overline{D}$ is positive if $\lambda = 0$. Note that the positivity of $\overline{D}$ implies that $\overline{D}^{1/2}$ exists and is positive. Given that Proposition 2.4.1 implies that $\overline{D} \in \mathcal{M}_{b,L}^{\hat{\zeta}}$, the positivity of $\overline{D}$ yields $\overline{D}^{1/2} \in \mathcal{M}_{b,L}^{\hat{\zeta}}$.

(v) The boundness of g , τ_1 , τ_2 and σ_2^0 (mentioned in (H1), (H3) and (H4)) is used in the proof of Proposition 2.3.7.

(vi) In some previous works on the problem without common noise (see [BS23b] and references therein), the coefficient of $\frac{1}{\epsilon}$ in the slow variable in (2.1.1) is chosen to be a function b , instead of a multiple of f , which appears in the fast variable as well. This choice will be explained in the last section.

(vii) If we were in a periodic setting (so instead of $\mathbb{R}^d$ we had $\mathbb{T}^d$), our Theorem 2.1.1 still holds under the same assumptions but without the inequality in (H2).

2.3.2 Preliminary results

A preliminary observation about system (2.1.1) is that the distribution of $(X_t^\epsilon, Y_t^\epsilon)$ in $\Omega^1 \times \Omega^0$, $t > 0$, remains in $\mathcal{A}$ provided that the distribution of $(X_0^\epsilon, Y_0^\epsilon)$ is in $\mathcal{A}$. We hence have the following proposition.

Proposition 2.3.7. *Suppose that (H1)-(H4) hold and $T > 0$. Let $(X_t^\epsilon, Y_t^\epsilon)$ be the solution of (2.1.1) with initial condition $(X_0, Y_0) = (\eta, \zeta) \sim m \in \mathcal{A}$; that is $\eta \in L^2(\Omega^1, \mathcal{F}^1, \mathbb{P}^1; \mathbb{R}^d)$ and $\zeta \in L^p(\Omega^1, \mathcal{F}^1, \mathbb{P}^1; \mathbb{R}^d)$ for every $p > 0$. Then, for any $p \in \mathbb{N}$ the following estimate holds*

$$\sup_{\epsilon \in (0,1]} \sup_{t \in [0,T]} \mathbb{E} \left[|Y_t^\epsilon|^{2p} \right] \leq C(p) + \mathbb{E}^1[|\zeta|^{2p}], \quad (2.3.1)$$

for some constant $C(p)$ depending on p . In particular, the conditional expectation

$\mathcal{L}(X_t^\epsilon, Y_t^\epsilon | \mathcal{F}_t^0)$ is in $\mathcal{A}$, $\mathbb{P}^0$ -a.s for any $\epsilon \in (0, 1]$ and any $t \in [0, T]$.

Proof. Fix $p \in \mathbb{N}$. We first note that since $\zeta \in L^{2p}(\Omega^1)$ we may construct $(X_t^\epsilon, Y_t^\epsilon)$ such that $\mathbb{E}[|Y_t^\epsilon|^{2p}]$ is finite for all $t \in [0, T]$. We only have to apply Itô's formula to $|Y_t^\epsilon|^{2p}$ and we take the expectation $\mathbb{E}$ to derive

$$\mathbb{E} \left[|Y_t^\epsilon|^{2p} \right] = \mathbb{E}[|\zeta|^{2p}] + \frac{2p}{\epsilon} \int_0^t \mathbb{E} \left[g \cdot Y_s^\epsilon |Y_s^\epsilon|^{2p-2} \right] ds + \frac{2p}{\epsilon^2} \mathbb{E} \left[\text{tr}(a) |Y_s^\epsilon|^{2p-2} \right] ds$$

$$+ \frac{2p}{\epsilon^2} \int_0^t \mathbb{E} \left[f \cdot Y_s^\epsilon |Y_s^\epsilon|^{2p-2} + 2(p-1) |Y_s^\epsilon|^{2p-2} a : Y_s^\epsilon \otimes Y_s^\epsilon \right] ds,$$

where a is the matrix defined at the introduction before relation (2.1.2). Since g and a are bounded (by Hypotheses (H1), (H3) and (H4)) and f satisfies (H2), the proof now follows as in Lemma 4.1 of [BS23b]. $\square$

2.4 The Poisson Equation

To prove Theorem 2.1.1 we will use some regularity results for the Poisson equation associated with the operator (2.1.15). In the following results ζ is a complete collection of multi-indices that contains all (n, l, β) with $|(n, l, \beta)| \leq 2$.

Proposition 2.4.1. *Assume (H1) and (H2) and suppose that $h \in \mathcal{M}_p^\zeta(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$, $f \in \mathcal{M}_p^\zeta(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$ and $a \in \mathcal{M}_p^\zeta(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$. Recall the definition of $L_0^{x, \mu}$ from (2.1.2) and of $\pi_{x, \mu}$ from (2.1.3). Then, $\bar{h}(x, \mu) := \int_{\mathbb{R}^d} h(x, y, \mu) \pi_{x, \mu}(dy) \in \mathcal{M}_{b, L}^\zeta(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$. Moreover, there exists a unique classical solution H to the problem*

$$\begin{cases} L_0^{x, \mu} H(x, y, \mu) = h(x, y, \mu) - \bar{h}(x, \mu), \\ \int_{\mathbb{R}^d} H(x, y, \mu) \pi_{x, \mu}(dy) = 0, \end{cases}$$

such that H has at most polynomial growth as $|y| \rightarrow \infty$ and the mixed derivatives $H, D_{y_i} H$ and $D_{y_i y_j}^2 H$ are in $\mathcal{M}_p^\zeta(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$ for every $i, j \in \{1, \dots, d\}$.

Proof. The first part of the statement follows from Lemma A.6 of [BS23b], while the second part is Lemma A.5 from the same paper. $\square$

Remark 2.4.2. By standard estimates for elliptic equations, the above Proposition also holds if $h \in \mathcal{M}_p^\zeta([0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$. In that case, the solution H is also in $\mathcal{M}_p^\zeta([0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$.

With the idea of doubling the variables in (2.1.15), we may consider the operator $L_{0,2}^{x,x',\mu}$ acting on $\phi \in C_b^2(\mathbb{R}^d \times \mathbb{R}^d)$ with

$$\begin{aligned} L_{0,2}^{x,x',\mu} \phi(y, y') &:= f(x, y, \mu) \cdot D_y \phi(y, y') + f(x', y', \mu) \cdot D_{y'} \phi(y, y') \\ &+ a(x, y, \mu) : D_{yy}^2 \phi(y, y') + a(x', y', \mu) : D_{y'y'}^2 \phi(y, y') \\ &+ \sigma_2^0(x, \mu) \sigma_2^{0\top}(x', \mu) : D_{yy'}^2 \phi(y, y'). \end{aligned} \quad (2.4.1)$$

We observe that due to assumption (H1), $L_{0,2}^{x,x',\mu}$ is non-degenerate elliptic. We also notice that the unique distributional solution $\Pi_{x,x',\mu}$ to the adjoint equation

$$\begin{cases} \left(L_{0,2}^{x,x',\mu} \right)^* \Pi_{x,x',\mu} = 0, \\ \int_{\mathbb{R}^{2d}} \Pi_{x,x',\mu}(dy, dy') = 1, \text{ for every } x, x' \in \mathbb{R}^d \text{ and } \mu \in \mathcal{P}_2(\mathbb{R}^d) \end{cases}$$

has y -marginal equal to $\pi_{x,\mu}$ and y' -marginal equal to $\pi_{x',\mu}$. Indeed, this follows by observing that $L_{0,2}^{x,x',\mu} \phi(y) = L_0^{x,\mu} \phi(y)$ and $L_{0,2}^{x,x',\mu} \phi(y) = L_0^{x',\mu} \phi(y')$, where $L_0^{x,\mu}$ was introduced in (2.1.2).

Lemma 2.4.3. *Assume Hypotheses (H1)-(H6). Let $\varphi : [0, T] \times \mathbb{R}^{4d} \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ be an element of $\mathcal{M}_p^\zeta([0, T] \times \mathbb{R}^{2d} \times \mathbb{R}^{2d} \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$. We also assume that*

$$\begin{aligned} \int_{\mathbb{R}^{2d}} \varphi(t, x, x', y, y', \mu) \Pi_{x,x',\mu}(dy, dy') &= 0, \\ \text{for every } t \in [0, T], x, x' \in \mathbb{R}^d, \mu \in \mathcal{P}_2(\mathbb{R}^d). \end{aligned} \quad (2.4.2)$$

Then, there exists a function $v : [0, T] \times \mathcal{A} \rightarrow \mathbb{R}$ such that for every $(t, m) \in [0, T] \times \mathcal{A}$

$$\mathcal{L}_0 v(t, m) = \int_{\mathbb{R}^{2d}} \int_{\mathbb{R}^{2d}} \varphi(t, x, x', y, y', \pi_*^1 m) m(dx, dy) m(dx', dy'). \quad (2.4.3)$$

Proof. Note that assumptions (H1)-(H6) guarantee that Proposition 2.4.1 and Remark 2.4.2

(for dimension $2d$) can be applied to the operator $L_{0,2}^{x,x',\mu}$, therefore there exists a function $\chi \in \mathcal{M}_p^\zeta([0, T] \times \mathbb{R}^{2d} \times \mathbb{R}^{2d} \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$ such that

$$L_0^{x,x',\mu} \chi(t, x, x', y, y', \mu) = \varphi(t, x, x', y, y', \mu). \quad (2.4.4)$$

We claim that $v : [0, T] \times \mathcal{A} \rightarrow \mathbb{R}$ with

$$v(t, m) = \int_{\mathbb{R}^{2d}} \int_{\mathbb{R}^{2d}} \chi(t, x, x', y, y', \pi_*^1 m) m(dx, dy) m(dx', dy') \quad (2.4.5)$$

satisfies (2.4.3). Indeed, since $\chi \in \mathcal{M}_p^\zeta$ we have

$$\begin{aligned} \frac{\delta v}{\delta m}(t, m, x, y) &= \int_{\mathbb{R}^{2d}} \chi(t, x, x', y, y', \pi_*^1 m) m(dx', dy') \\ &+ \int_{\mathbb{R}^{2d}} \chi(t, x', x, y', y, \pi_*^1 m) m(dx', dy') \\ &+ \int_{\mathbb{R}^{4d}} \frac{\delta \chi}{\delta m}(t, \bar{x}, x', \bar{y}, y', \pi_*^1 m, x) m(d\bar{x}, d\bar{y}) m(dx', dy'). \end{aligned} \quad (2.4.6)$$

In addition,

$$\begin{aligned} \frac{\delta^2 v}{\delta m^2}(t, m, x, y, x', y') &= \chi(t, x, x', y, y', \pi_*^1 m) + \chi(t, x', x, y', y, \pi_*^1 m) \\ &+ \int_{\mathbb{R}^{2d}} \frac{\delta \chi}{\delta m}(t, x, \bar{x}, y, \bar{y}, \pi_*^1 m, x') m(d\bar{x}, d\bar{y}) \\ &+ \int_{\mathbb{R}^{2d}} \frac{\delta \chi}{\delta m}(t, \bar{x}, x, \bar{y}, y, \pi_*^1 m, x') m(d\bar{x}, d\bar{y}) \\ &+ \int_{\mathbb{R}^{4d}} \frac{\delta^2 \chi}{\delta m^2}(t, \bar{x}, \tilde{x}, \bar{y}, \tilde{y}, \pi_*^1 m, x, x') m(d\bar{x}, d\bar{y}) m(d\tilde{x}, d\tilde{y}) \\ &+ \int_{\mathbb{R}^{4d}} \frac{\delta \chi}{\delta m}(t, \bar{x}, x', \bar{y}, y', \pi_*^1 m, x) m(d\bar{x}, d\bar{y}) \\ &+ \int_{\mathbb{R}^{4d}} \frac{\delta \chi}{\delta m}(t, x', \bar{x}, y', \bar{y}, \pi_*^1 m, x) m(d\bar{x}, d\bar{y}). \end{aligned} \quad (2.4.7)$$

Notice that only the first two terms depend on both y, y' . Therefore,

$$\begin{aligned}
\mathcal{L}_0 v(t, m) &= \int_{\mathbb{R}^{2d}} \int_{\mathbb{R}^{2d}} L_0^{x, \mu} \chi(t, x, x', y, y', \pi_*^1 m) m(dx', dy') m(dx, dy) \\
&+ \int_{\mathbb{R}^{4d}} \frac{\sigma^0(x, \pi_*^1 m) \sigma^{0\top}(x', \pi_*^1 m)}{2} : D_{yy'}^2 \chi(t, x', x, y', y, \pi_*^1 m) m(dx, dy) m(dx', dy') \\
&+ \int_{\mathbb{R}^{4d}} \frac{\sigma^0(x, \pi_*^1 m) \sigma^{0\top}(x', \pi_*^1 m)}{2} : D_{yy'}^2 \chi(t, x, x', y, y', \pi_*^1 m) m(dx, dy) m(dx', dy') \\
&+ \int_{\mathbb{R}^{2d}} \int_{\mathbb{R}^{2d}} L_0^{x, \mu} \chi(t, x', x, y', y, \pi_*^1 m) m(dx', dy') m(dx, dy). \tag{2.4.8}
\end{aligned}$$

We observe that, due to $D_{yy'}^2 \chi(t, x', x, y', y, \pi_*^1 m)$ being a symmetric matrix, the fourth term is equal to

$$\begin{aligned}
&\int_{\mathbb{R}^{4d}} \frac{\sigma^0(x', \pi_*^1 m) \sigma^{0\top}(x, \pi_*^1 m)}{2} : D_{yy'}^2 \chi(t, x', x, y', y, \pi_*^1 m) m(dx, dy) m(dx', dy') = \\
&\int_{\mathbb{R}^{4d}} \frac{\sigma^0(x, \pi_*^1 m) \sigma^{0\top}(x', \pi_*^1 m)}{2} : D_{yy'}^2 \chi(t, x, x', y, y', \pi_*^1 m) m(dx', dy') m(dx, dy),
\end{aligned}$$

which is equal to the third term. Also, after renaming the variables, the second term is equal to $\int_{\mathbb{R}^{2d}} \int_{\mathbb{R}^{2d}} L_0^{x', \mu} \chi(t, x, x', y, y', \pi_*^1 m) m(dx', dy') m(dx, dy)$. Thus, going back to (2.4.8),

$$\mathcal{L}_0 v(t, m) = \int_{\mathbb{R}^{4d}} L_{0,2}^{x, x', \mu} \chi(t, x, x', y, y', \mu) m(dx, dy) m(dx', dy').$$

Combining this with (2.4.4) yields (2.4.3). $\square$

Finally, we prove the following lemma which will provide us with useful estimates when we introduce the correctors u_1, u_2 .

Lemma 2.4.4. *Assume that Hypotheses (H1)-(H6) hold and let $\chi : [0, T] \times \mathbb{R}^{4d} \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ be an element of $\mathcal{M}_p^{\zeta}([0, T] \times \mathbb{R}^{2d} \times \mathbb{R}^{2d} \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$. We define the function $v : [0, T] \times \mathcal{A} \rightarrow \mathbb{R}$ as in (2.4.5). Then for every $m \in \mathcal{A}$ there exists a constant C depending only on $T, \pi_*^2 m$*

and χ such that

$$\sup_{t \in [0, T]} \left\{ |v(t, m)| + |\dot{v}(t, m)| + |\mathcal{L}_1 v(t, m)| + |\mathcal{L}_2 v(t, m)| \right\} \leq C, \quad (2.4.9)$$

where $\mathcal{L}_1, \mathcal{L}_2$ are as in (2.1.14) and (2.1.13).

Proof. We essentially have to prove four estimates. For any $(t, m) \in [0, T] \times \mathcal{A}$ we have, by the definition of the space $\mathcal{M}_p^\zeta$ and Remark 2.3.5

$$\begin{aligned} |v(t, m)| &\leq C \int_{\mathbb{R}^{4d}} (1 + |y|^{p_\chi} + |y'|^{p_\chi}) m(dx, dy) m(dx', dy') \\ &\leq C \left(1 + \int_{\mathbb{R}^d} |y|^{p_\chi} \pi_*^2 m(dy) \right), \end{aligned} \quad (2.4.10)$$

and likewise

$$\begin{aligned} |\dot{v}(t, m)| &\leq \int_{\mathbb{R}^{4d}} |\dot{\chi}(t, x, x', y, y', \pi_*^1 m)| m(dx, dy) m(dx', dy') \\ &\leq C \left(1 + \int_{\mathbb{R}^d} |y|^{p_\chi} \pi_*^2 m(dy) \right). \end{aligned} \quad (2.4.11)$$

To bound $|\mathcal{L}_1 v|$ and $|\mathcal{L}_2 v|$ we will use (2.4.6), (2.4.7). It is easy to see that after plugging (2.4.6), (2.4.7) in $\mathcal{L}_1 v$ and $\mathcal{L}_2 v$, as given by (2.1.14) and (2.1.13), we get two sums of terms of the form $\int_{\mathbb{R}^{4d}} \phi(t, x, x', y, y', \pi_*^1 m) m(dx, dy) m(dx', dy')$ with ϕ being a function depending linearly on the derivatives of χ and the coefficients of (2.1.1). Since all these functions are in some $\mathcal{M}_p$, we deduce that any such ϕ is in some $\mathcal{M}_p^{\zeta'}$ (here ζ' is a set of multi-indices) and hence we can bound any term

$$\left| \int_{\mathbb{R}^{4d}} \phi(t, x, x', y, y', \pi_*^1 m) m(dx, dy) m(dx', dy') \right| \leq C \left(1 + \int_{\mathbb{R}^d} |y|^p \pi_*^2 m(dy) \right)$$

for some p depending on χ and the coefficients of (2.1.1). Thus, for $(t, m) \in [0, T] \times \mathcal{A}$ we

deduce the bound

$$|\mathcal{L}_1 v(t, m)| + |\mathcal{L}_2 v(t, m)| \leq C \left(1 + \int_{\mathbb{R}^d} |y|^p \pi_*^2 m(dy) \right), \quad (2.4.12)$$

for some $p \in \mathbb{N}$. Since $m \in \mathcal{A}$ ($\pi_*^2 m$ finite moments), we combine (2.4.10), (2.4.11) and (2.4.12) and (2.4.9) follows. $\square$

Remark 2.4.5. Due to the linear dependence of all the terms of v , $\dot{v}$, $\mathcal{L}_1 v$ and $\mathcal{L}_2 v$ on χ , it follows from the proof that the constant in (2.4.9) is a multiple of C_χ , where C_χ is defined in Remark 2.3.5.

2.5 Cauchy problem associated to McKean-Vlasov processes

As mentioned in the introduction we consider the McKean-Vlasov control equation associated with (2.1.7) which is (2.1.9). The following theorem holds.

Theorem 2.5.1. *Assume Hypotheses (H1)-(H6) and that $G \in \mathcal{M}_{b,L}^{\dot{\zeta}}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$, where $\dot{\zeta}$ is as in Theorem 2.1.1. Then, there exists a unique solution $u \in \mathcal{M}_{b,L}^{\dot{\zeta}}([0, T] \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$ of (2.1.9) with*

$$\|u\|_{\mathcal{M}_b^{\dot{\zeta}}} \leq C(T) \|G\|_{\mathcal{M}_b^{\dot{\zeta}}}, \quad (2.5.1)$$

for some constant $C(T)$ that depends only on T , the norms of the coefficients of (2.1.9):

$\|\bar{\gamma}\|_{\mathcal{M}_b^{\hat{\zeta}}}$, $\|\bar{D}^{1/2}\|_{\mathcal{M}_b^{\hat{\zeta}}}$ and their Lipschitz constants.

Furthermore $u(t, \mu) = \mathbb{E}^0[G(\mathcal{L}(X_T^{\mu,t} | \mathcal{F}_T^0))]$ and the following estimate holds

$$\sup_{\substack{t \in [0, T], \mu \in \mathcal{P}_2(\mathbb{R}^d) \\ x_1, x_2 \in \mathbb{R}^d}} \left\{ |D_\mu \dot{u}(t, \mu, x_1)| + |D_{x_1} D_\mu \dot{u}(t, \mu, x_1)| \right. \\ \left. + |D_{\mu\mu}^2 \dot{u}(t, \mu, x_1, x_2)| \right\} \leq C(T) \|G\|_{\mathcal{M}_b^{\dot{\zeta}}}. \quad (2.5.2)$$

Proof. First we observe that $\bar{\gamma}, \bar{D} \in \mathcal{M}_p^{\hat{\mathcal{S}}}$ follows from Proposition 2.4.1. In addition, since $\bar{D}$ is positive (assumption (H6)) and the square root map is Fréchet differentiable up to arbitrary order with all derivatives being bounded on sets of uniformly positive matrices, we also have that $\bar{D}^{1/2} \in \mathcal{M}_p^{\hat{\mathcal{S}}}$ (see also [CLX22] for a more detailed discussion). Now, the result follows from Theorem 2.7.1. $\square$

2.6 Proof of the main result

In this section we will prove the main result of this paper. For motivation purposes, we will first show a heuristic calculation which sketches the proof of the convergence described in section 1.2. We will be using the notation $\mu = \pi_*^1 m$.

2.6.1 Heuristic calculation

We start by making the ansatz $u^\epsilon(t, m) = u(t, m) + \epsilon u_1(t, m) + \epsilon^2 u_2(t, m)$ in (2.1.12) and we collect the $\mathcal{O}(\frac{1}{\epsilon^2})$, $\mathcal{O}(\frac{1}{\epsilon})$ and $\mathcal{O}(1)$ terms. Consequently, we write

$$\frac{\mathcal{L}_0 u}{\epsilon^2} + \frac{\mathcal{L}_1 u + \mathcal{L}_0 u_1}{\epsilon} + \mathcal{L}_0 u_2 + \mathcal{L}_1 u_1 + \mathcal{L}_2 u - \mathcal{L}u + \dot{u} + \mathcal{L}u = \mathcal{O}(\epsilon),$$

where $\mathcal{L}, \mathcal{L}_0, \mathcal{L}_1, \mathcal{L}_2$ are defined in (2.1.10), (2.1.15), (2.1.14), (2.1.13), respectively. We want the following relations to hold

$$\mathcal{O}\left(\frac{1}{\epsilon^2}\right) : \quad \mathcal{L}_0 u = 0, \tag{2.6.1}$$

$$\mathcal{O}\left(\frac{1}{\epsilon}\right) : \quad \mathcal{L}_0 u_1 + \mathcal{L}_1 u = 0, \tag{2.6.2}$$

$$\mathcal{O}(1) : \quad \mathcal{L}_0 u_2 = -\mathcal{L}_1 u_1 - \mathcal{L}_2 u + \mathcal{L}u. \tag{2.6.3}$$

The first equation is satisfied when $\frac{\delta u}{\delta m}(t, m, x, y)$ is independent of y , so by Proposition 2.2.6, $u(t, m) = u(t, \pi_*^1 m)$; that is, $u(\cdot, m)$ depends only on $\pi_*^1 m$. The second equation, can

be written as

$$\mathcal{L}_0 u_1(t, m) = -\lambda \int_{\mathbb{R}^{2d}} \tilde{f}(x, y, m) \cdot D_m u(t, \pi_*^1 m, x) m(dx, dy),$$

so we expect $u_1(t, m) = -\lambda \int_{\mathbb{R}^{2d}} y \cdot D_m u(t, \mu, x) m(dx, dy)$. Now, due to Lemma 2.4.3 there exists a solution u_2 of (2.6.3). Putting everything together, we deduce that u satisfies

$$\dot{u}(t, \mu) + \mathcal{L}u(t, \mu) = \mathcal{O}(\epsilon),$$

thus, letting $\epsilon \rightarrow 0$, we deduce that u^ϵ converges to u , which satisfies the equation in (2.1.9).

2.6.2 Proof of the main theorem

We will construct the corrector u_1 first.

Proposition 2.6.1. *Assume Hypotheses (H1)-(H6) and let $\bar{u}$ be the classical solution of (2.1.9) given by Theorem 2.5.1. We also consider the function $u : [0, T] \times \mathcal{P}_2(\mathbb{R}^{2d}) \rightarrow \mathbb{R}$ with $u(t, m) := \bar{u}(t, \pi_*^1 m)$. Then, there exists a function $u_1 : [0, T] \times \mathcal{A} \rightarrow \mathbb{R}$ satisfying*

$$\mathcal{L}_0 u_1(t, m) = -\mathcal{L}_1 u(t, m) \tag{2.6.4}$$

in the classical sense. Moreover, for every $m \in \mathcal{A}$ there exists a constant $C > 0$ depending only on T and $\pi_*^2 m$ such that

$$\sup_{t \in [0, T]} \left\{ |u_1(t, m)| + |\dot{u}_1(t, m)| + |\mathcal{L}_2 u_1(t, m)| \right\} \leq C \|G\|_{\mathcal{M}_{b, L}^\xi}. \tag{2.6.5}$$

Proof. First, we note that since $u(t, m)$ depends only on the x -marginal of m , by Proposition 2.2.5 $\frac{\delta u}{\delta m}(t, m, x)$ is independent of y , therefore $\mathcal{L}_1 u = \lambda \int_{\mathbb{R}^{2d}} f(x, y, \mu) \cdot D_\mu \bar{u}(t, \mu, x) m(dx, dy)$.

For $(t, m) \in [0, T] \times \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ we define u_1 to be such that

$$u_1(t, m) = -\lambda \int_{\mathbb{R}^{2d}} y \cdot D_\mu \bar{u}(t, \pi_*^1 m, x) m(dx, dy).$$

By [CD⁺18a] (section 5.2.2 Example 3) and Proposition 2.2.5, since the dependence of the integrand with respect to the measure is on $\pi_*^1 m$, we have

$$\frac{\delta u_1}{\delta m}(t, m, x, y) = -\lambda y \cdot D_\mu \bar{u}(t, \pi_*^1 m, x) - \lambda \int_{\mathbb{R}^{2d}} y' \cdot \frac{\delta(D_\mu \bar{u})}{\delta \mu}(t, \pi_*^1 m, x', x) m(dx', dy'),$$

while $\frac{\delta^2 u_1}{\delta m^2}(t, m, x, y, x', y')$ has no term depending on both y and y' . So, using (2.1.15), we get $\mathcal{L}_0 u_1 = -\lambda \int_{\mathbb{R}^{2d}} f(x, y, \pi_*^1 m) \cdot D_\mu \bar{u}(t, \pi_*^1 m, x) m(dx, dy) = -\mathcal{L}_1 u(t, m)$, which is (2.6.4). To prove (2.6.5) we will apply Lemma 2.4.4 for $\chi_1(t, x, x', y, y', \mu) := y \cdot D_\mu \bar{u}(t, \mu, x)$; that is χ_1 is constant in x', y' . We observe by Theorem 2.5.1 that $\chi_1 \in \mathcal{M}_p^{\zeta_1}([0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$, where ζ_1 is a set of multi-indices containing all (n, l, β) with $|(n, l, \beta)| \leq 3$, as indicated by Theorem 2.5.1. In addition, by (2.5.1), $C_{\chi_1} = C(T) \|G\|_{\mathcal{M}_p^{\dot{\zeta}}}$. Having this in mind, estimate (2.6.5) follows from Lemma 2.4.4 and Remark 2.4.5. $\square$

Next, we will construct the second corrector u_2 .

Proposition 2.6.2. *Suppose that hypotheses (H1)-(H6) hold and let $u, \bar{u}, u_1$ be as in Proposition 2.6.1. Then, there exists a function $u_2 : [0, T] \times \mathcal{A} \rightarrow \mathbb{R}$ such that*

$$\mathcal{L}_0 u_2(t, m) = -\mathcal{L}_1 u_1 - \mathcal{L}_2 u + \mathcal{L} u. \quad (2.6.6)$$

Furthermore, for every $m \in \mathcal{A}$ there exists a constant $C > 0$ depending only on T and $\pi_*^2 m$ such that

$$\sup_{\substack{t \in [0, T], \\ m \in \mathcal{A}}} \left\{ |u_2(t, m)| + |\dot{u}_2(t, m)| + |\mathcal{L}_1 u_2(t, m)| + |\mathcal{L}_2 u_2(t, m)| \right\} \leq C \|G\|_{\mathcal{M}_{b,L}^{\dot{\zeta}}}. \quad (2.6.7)$$

Proof. After plugging in (2.1.14), (2.1.13), (2.1.10) and the formula for u_1 in the right hand side of (2.6.6) a straightforward simplification yields that it is equal to

$$\int_{\mathbb{R}^{2d}} (\bar{c}(x, \mu) - c(x, y, \mu)) \cdot D_\mu \bar{u}(t, \mu, x) m(dx, dy) \quad (2.6.8)$$

$$+ \lambda \int_{\mathbb{R}^{2d}} (g(x, y, \mu) - \bar{g}(x, \mu)) \cdot D_\mu \bar{u}(t, \mu, x) m(dx, dy) \quad (2.6.9)$$

$$+ \int_{\mathbb{R}^{2d}} \left(\frac{\overline{\sigma \sigma^\top}(x, \mu) - \sigma \sigma^\top(x, y, \mu)}{2} \right) : D_x D_\mu \bar{u}(t, \mu, x) m(dx, dy) \quad (2.6.10)$$

$$+ \lambda^2 \int_{\mathbb{R}^{2d}} \left(\frac{y \otimes f(x, y, \mu) + (y \otimes f(x, y, \mu))^\top - \overline{y \otimes f(x, y, \mu) + (y \otimes f(x, y, \mu))^\top}}{2} \right) : D_x D_\mu \bar{u}(t, \mu, x) m(dx, dy) \quad (2.6.11)$$

$$+ \lambda \int_{\mathbb{R}^{2d}} \left(\frac{\tau_1 \sigma^\top + \sigma \tau_1^\top}{2}(x, y, \mu) - \frac{\overline{\tau_1 \sigma^\top + \sigma \tau_1^\top}}{2}(x, \mu) \right) : D_x D_\mu \bar{u}(t, \mu, x) m(dx, dy) \quad (2.6.12)$$

$$+ \lambda^2 \int_{\mathbb{R}^{2d}} \int_{\mathbb{R}^{2d}} f(x, y, \mu) \cdot D_{\mu\mu}^2 \bar{u}(t, \mu, x, x') y' m(dx', dy') m(dx, dy) \quad (2.6.13)$$

$$+ \frac{\lambda^2}{2} \int_{\mathbb{R}^{2d}} \int_{\mathbb{R}^{2d}} \sigma_2^0(x, \mu) \sigma_2^{0\top}(x', \mu) : D_{\mu\mu}^2 \bar{u}(t, \mu, x, x') \mu(dx) \mu(dx'). \quad (2.6.14)$$

We note that all the integrands satisfy (2.4.2) except for the last two. Hence, by Proposition 2.4.1, there exist $u_{i,2} \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$, $i \in \{1, 2\}$ and $u_{i,2} \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$, $i = 3, 4, 5$ such that

$$L_{0,2}^{x,x',\mu} u_{2,1} = \bar{c}(x, \mu) - c(x, y, \mu) \quad (2.6.15)$$

$$L_{0,2}^{x,x',\mu} u_{2,2} = \lambda(g(x, y, \mu) - \bar{g}(x, \mu)) \quad (2.6.16)$$

$$L_{0,2}^{x,x',\mu} u_{2,3} = \lambda^2 \frac{y \otimes f(x, y, \mu) + (y \otimes f(x, y, \mu))^\top}{2} - \lambda^2 \frac{\overline{y \otimes f(x, y, \mu) + (y \otimes f(x, y, \mu))^\top}}{2} \quad (2.6.17)$$

$$L_{0,2}^{x,x',\mu} u_{2,4} = \frac{\overline{\sigma \sigma^\top}(x, \mu) - \sigma \sigma^\top(x, y, \mu)}{2} \quad (2.6.18)$$

$$L_{0,2}^{x,x',\mu} u_{2,5} = \lambda \frac{\tau_1 \sigma^\top + \sigma \tau_1^\top}{2}(x, y, \mu) - \lambda \frac{\overline{\tau_1 \sigma^\top + \sigma \tau_1^\top}}{2}(x, \mu). \quad (2.6.19)$$

We also, have for $u_{2,6}(t, \mu, x, x', y, y') = \frac{\lambda^2}{2} y \cdot D_{\mu\mu}^2 \bar{u}(t, \mu, x, x') y'$

$$\begin{aligned} L_{0,2}^{x,x',\mu} u_{2,6} &= \frac{\lambda^2}{2} f(x, y, \mu) \cdot D_{\mu\mu}^2 u(t, \mu, x, x') y' + \frac{\lambda^2}{2} y \cdot D_{\mu\mu}^2 u(t, \mu, x, x') f(x', y', \mu) \\ &\quad + \frac{\lambda^2}{2} \sigma_2^0(x, \mu) \sigma_2^{0\top}(x', \mu) : D_{\mu\mu}^2 \bar{u}(t, \mu, x, x'). \end{aligned}$$

Note that from Theorem 2.5.1, it follows that $u_{2,6} \in \mathcal{M}_p^\zeta([0, T] \mathbb{R}^{2d} \times \mathbb{R}^{2d} \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$ with ζ as described in section 3. In addition, since $D_{\mu\mu}^2 u$ is symmetric with respect to x, x' (Lemma 2.4 from [CDLL19]), we deduce that $\int_{\mathbb{R}^{4d}} L_{0,2}^{x,x',\mu} u_{2,6} m(dx, dy) m(dx', dy')$ is equal to the sum of the quantities (2.6.13) and (2.6.14). For $(t, m) \in [0, T] \times \mathcal{A}$ we now define

$$\begin{aligned} u_2(t, m) &:= \int_{\mathbb{R}^{4d}} u_{2,6}(t, \mu, x, x', y, y') m(dx, dy) m(dx', dy') \\ &\quad + \sum_{i=1}^2 \int_{\mathbb{R}^{2d}} u_{2,i}(\mu, x, y) \cdot D_\mu \bar{u}(t, \mu, x) m(dx, dy) \\ &\quad + \sum_{i=3}^5 \int_{\mathbb{R}^{2d}} u_{2,i}(\mu, x, y) : D_x D_\mu \bar{u}(t, \mu, x) m(dx, dy). \end{aligned} \tag{2.6.20}$$

By repeating the calculation from Lemma 2.4.3, by the construction of $u_{2,i}$, $i \in \{1, 2, \dots, 6\}$ we deduce that $\mathcal{L}_0 u_2$ is equal to the sum of the quantites (2.6.8)-(2.6.14). Therefore, we derive (2.6.6).

The estimate (2.6.7), follows from Lemma 2.4.4. Indeed, due to the regularity of the $u_{2,i}$, $i = 1, 2, \dots, 6$ and Theorem 2.5.1, Lemma 2.4.4 can be applied to every term of u_2 . The quantity $C \|G\|_{\mathcal{M}_{b,L}^\zeta}$ appears on the right hand side of (2.6.7) as in the proof of Proposition 2.6.1 because of Remark 2.4.5. $\square$

We are now ready to prove Theorem 2.1.1

Proof. We will make the heuristic argument provided in the previous subsection rigorous.

Now let u be the unique classical solution of (2.1.9)³. For $t, m \in [0, T] \times \mathcal{P}_2(\mathbb{R}^{2d})$, we consider

$$v^\epsilon(t, m) = u(t, \pi_*^1 m) + \epsilon u_1(t, m) + \epsilon^2 u_2(t, m), \quad (2.6.21)$$

where u_1, u_2 are given by Propositions 2.6.1 and 2.6.2. We will exploit the fact that v^ϵ is “almost” a solution to (2.1.12), as the heuristic calculation in section 5.1 indicates.

Let $m \in \mathcal{A}$ and $t \in [0, T]$. For $s \in [t, T]$, we denote by m_s the conditional joint law $\mathcal{L}((X_s^\epsilon, Y_s^\epsilon)^{t, m} | \mathcal{F}_s^0)$ of the pair $(X_t^\epsilon, Y_t^\epsilon)$ satisfying (2.1.1) given that the initial distribution (at $s = t$) is m , and we also set $\mu_s = \pi_*^1 m_s$. Note that since $m \in \mathcal{A}$, by Proposition 2.3.7, we deduce that $m_s \in \mathcal{A}$, $\mathbb{P}^0$ -a.s for any $s \in [t, T]$ and $\epsilon \in (0, 1)$. We observe that by the construction of v^ϵ and Itô’s formula (Theorem 4.14 of [CD⁺18c]) we have

$$\begin{aligned} \mathbb{E}^0[v^\epsilon(T, m_T)] &= \mathbb{E}^0[v^\epsilon(t, m)] + \mathbb{E}^0\left[\int_t^T \left(\dot{v}^\epsilon(s, m_s) + \mathcal{L}^\epsilon v^\epsilon(s, m_s)\right) ds\right] \\ &= \mathbb{E}^0(v^\epsilon(t, m)) + \mathbb{E}^0\left[\int_t^T \left(\dot{u}(s, \mu_s) + \mathcal{L}u(s, \mu_s)\right) ds\right] \\ &\quad + \mathbb{E}^0\left[\int_t^T (\epsilon \dot{u}_1(s, m_s) + \epsilon^2 \dot{u}_2(s, m_s)) ds\right] \\ &\quad + \mathbb{E}^0\left[\int_t^T \left(\epsilon^2 \mathcal{L}_2 u_2(s, m_s) + \epsilon \mathcal{L}_1 u_2(s, m_s) + \epsilon \mathcal{L}_2 u_1(s, m_s)\right) ds\right]. \end{aligned}$$

Hence, since u satisfies (2.1.9),

$$\begin{aligned} \mathbb{E}^0[v^\epsilon(T, m_T) - v^\epsilon(t, m)] \\ = \mathbb{E}^0\left[\int_t^T \left(\epsilon \dot{u}_1 + \epsilon^2 \dot{u}_2 + \epsilon^2 \mathcal{L}_2 u_2 + \epsilon \mathcal{L}_1 u_2 + \epsilon \mathcal{L}_2 u_1\right)(s, m_s) ds\right]. \end{aligned} \quad (2.6.22)$$

3. In comparison with the previous propositions, we drop the bar for simplicity.

Now, by Proposition 2.3.7, $m_s \in \mathcal{A}$, so using the estimates (2.6.5) and (2.6.6), we deduce

$$\left| \mathbb{E}^0 \left[v^\epsilon(T, m_T) - v^\epsilon(t, m) \right] \right| \leq \epsilon C(T) \|G\|_{\mathcal{M}_b^\zeta}, \quad (2.6.23)$$

where $C(T)$ is some constant depending on T and $\pi_*^2 m$. Finally, it follows from the definition of u , v^ϵ and the triangle inequality that

$$\begin{aligned} \left| \mathbb{E}^0 \left[G(\mu_T) - G(\mathcal{L}(X_T^{t,\mu} | \mathcal{F}_T^0)) \right] \right| &= \left| \mathbb{E}^0 \left[u(T, \mu_T) - u(t, \mu) \right] \right| \\ &\leq \left| \mathbb{E}^0 \left[v^\epsilon(T, m_T) - v^\epsilon(t, m) \right] \right| + \left| \mathbb{E}^0 \left[\epsilon u_1(T, m_T) - \epsilon u_1(t, m) + \epsilon^2 u_2(T, m_T) - \epsilon^2 u_2(t, m) \right] \right| \\ &\leq \epsilon C(T) \|G\|_{\mathcal{M}_b^\zeta}, \end{aligned}$$

where in the last inequality we used (2.6.23), (2.6.5) and (2.6.7). Thus, we derive (2.1.11).

The proof is complete. $\square$

2.7 On the McKean-Vlasov control equation

Let the maps B , Σ^1 , Σ^0 be such that $B \in \mathcal{M}_{b,L}^{\hat{\zeta}}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$ and $\Sigma^1, \Sigma^0 \in \mathcal{M}_{b,L}^{\hat{\zeta}}(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^{d \times d})$, where $\hat{\zeta}$ is as introduced at the beginning of section 2.1. In the probability spaces where (2.1.1) is posed we consider for $s \in [0, T]$, $t \in [s, T]$ the conditional McKean-Vlasov SDE:

$$\begin{aligned} X_t^{s,\mu} &= \eta + \int_s^t B(X_\tau^{s,\mu}, \mathcal{L}(X_\tau^{s,\mu} | \mathcal{F}_\tau^0)) d\tau + \int_s^t \Sigma^1(X_\tau^{s,\mu}, \mathcal{L}(X_\tau^{s,\mu} | \mathcal{F}_\tau^0)) dW_\tau \\ &\quad + \int_s^t \Sigma^0(X_\tau^{s,\mu}, \mathcal{L}(X_\tau^{s,\mu} | \mathcal{F}_\tau^0)) dB_\tau^0, \end{aligned} \quad (2.7.1)$$

where $\mathcal{L}(\eta) = \mu$. For $x \in \mathbb{R}^d$ we also consider the process $X_t^{s,x,\mu}$, which is such that

$$X_t^{s,x,\mu} = x + \int_s^t B(X_\tau^{s,x,\mu}, \mathcal{L}(X_\tau^{s,\mu} | \mathcal{F}_\tau^0)) d\tau + \int_s^t \Sigma^1(X_\tau^{s,x,\mu}, \mathcal{L}(X_\tau^{s,\mu} | \mathcal{F}_\tau^0)) dW_\tau$$

$$+ \int_s^t \Sigma^0(X_t^{s,x,\mu}, \mathcal{L}(X_\tau^{s,\mu} | \mathcal{F}_\tau^0)) dB_\tau^0. \quad (2.7.2)$$

Note that $X_t^{s,x,\mathcal{L}(\eta)} \Big|_{x=\eta} = X_t^{s,\mathcal{L}(\eta)}$. In addition, by standard estimates for conditional McKean-Vlasov SDEs proved in [CD⁺18c], which are consequences of Grönwall's and Burkholder-Davis-Gundy's (BDG for short) inequalities, we have that for any $p \geq 2$ there exists a constant $C_p > 0$ such that

$$\mathbb{E} \left[\sup_{t \in [s, T]} |X_t^{s,\mu} - X_t^{s,\mu'}|^p \right] \leq C_p \mathbf{d}_2(\mu, \mu')^p, \quad (2.7.3)$$

$$\mathbb{E}^0 \left[\sup_{t \in [s, T]} \mathbf{d}_2 \left(\mathcal{L}(X_t^{s,\mu} | \mathcal{F}_t^0), \mathcal{L}(X_t^{s,\mu'} | \mathcal{F}_t^0) \right)^p \right] \leq C_p \mathbf{d}_2(\mu, \mu')^p, \quad (2.7.4)$$

$$\mathbb{E} \left[\sup_{t \in [s, T]} |X_t^{s,x,\mu} - X_t^{s,x',\mu'}|^p \right] \leq C_p (|x - x'|^p + \mathbf{d}_2(\mu, \mu')^p) \quad (2.7.5)$$

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

Notation (i) For $\eta \in L^2(\mathcal{F}_s)$ with $\mathcal{L}(\eta) = \mu$, since we have estimate (2.7.5) we may use the notation $X_t^{s,x,\mu} = X_t^{s,x,\mathcal{L}(\eta)} = X_t^{s,x,\eta}$ and $X_t^{s,x,\mu} \Big|_{x=\eta} = X_t^{s,\mu} = X_t^{s,\mathcal{L}(\eta)}$.

(ii) Let $(\widetilde{\Omega}^1, \widetilde{\mathcal{F}}^1, \widetilde{\mathbb{P}}^1)$ be a copy of $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$. Any random variable X over $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$ is represented by a pointwise copy $\widetilde{X}$ on $(\widetilde{\Omega}^1, \widetilde{\mathcal{F}}^1, \widetilde{\mathbb{P}}^1)$.

In this section we will prove the following Theorem:

Theorem 2.7.1. *Let B, Σ^1, Σ^0 be as above and $G \in \mathcal{M}_{b,L}^{\dot{\boldsymbol{\zeta}}}(\mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$, where $\dot{\boldsymbol{\zeta}}$ is as in Theorem 2.1.1. Then, the function $v : [0, T] \times \mathcal{P}_2(\mathbb{R}^d) \rightarrow \mathbb{R}$ with*

$$v(t, \mu) = \mathbb{E}^0 \left[G(\mathcal{L}(X_T^{t,\mu} | \mathcal{F}_T^0)) \right]$$

is in $\mathcal{M}_{b,L}^{\dot{\boldsymbol{\zeta}}}([0, T] \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R})$, satisfies $\|v\|_{\mathcal{M}_b^{\dot{\boldsymbol{\zeta}}}} \leq C \|G\|_{\mathcal{M}_b^{\dot{\boldsymbol{\zeta}}}}$ for some constant $C > 0$

depending on T and is the unique solution to the PDE

$$\left\{ \begin{array}{l} \dot{v}(t, \mu) + \int_{\mathbb{R}^d} \left(B(x, \mu) \cdot D_\mu v(t, \mu, x) + \frac{1}{2} (\Sigma^1 \Sigma^{1\top} + \Sigma^0 \Sigma^{0\top})(s, \mu) : D_x D_\mu v(t, \mu, x) \right) \mu(dx) \\ \quad + \frac{1}{2} \int_{\mathbb{R}^{2d}} \Sigma^0(x, \mu) \Sigma^{0\top}(x', \mu) : D_{\mu\mu}^2 v(t, \mu, x, x') \mu(dx) \mu(dx') = 0, \\ \quad \quad \quad (t, \mu) \in [0, T) \times \mathcal{P}_2(\mathbb{R}^d), \\ v(T, \mu) = G(\mu), \quad \mu \in \mathcal{P}_2(\mathbb{R}^d). \end{array} \right. \quad (2.7.6)$$

Furthermore, the following estimate holds

$$\sup_{\substack{t \in [0, T], \mu \in \mathcal{P}_2(\mathbb{R}^d) \\ x_1, x_2 \in \mathbb{R}^d}} \left\{ |D_\mu \dot{v}(t, \mu, x_1)| + |D_{x_1} D_\mu \dot{v}(t, \mu, x_1)| + |D_{\mu\mu}^2 \dot{v}(t, \mu, x_1, x_2)| \right\} \leq C(T) \|G\|_{\mathcal{M}_b^\xi}. \quad (2.7.7)$$

The proof is inspired by the ideas presented in [BDL17]. For simplicity, we assume that $d = 1$ as the argument can be easily extended for any dimension d .

Proposition 2.7.2. *Let $X_t^{s, x, \mu}$ be as in (2.7.2). Then, $D_x X_t^{s, x, \mu}$ exists and is the unique solution of the SDE*

$$\begin{aligned} D_x X_t^{s, x, \mu} &= 1 + \int_s^t D_x B(X_\tau^{s, x, \mu}, \mu_\tau) D_x X_\tau^{s, x, \mu} d\tau + \int_s^t D_x \Sigma^1(X_\tau^{s, x, \mu}, \mu_\tau) D_x X_\tau^{s, x, \mu} dW_\tau \\ &\quad + \int_s^t D_x \Sigma^0(X_\tau^{s, x, \mu}, \mu_\tau) D_x X_\tau^{s, x, \mu} dB_\tau^0, \end{aligned} \quad (2.7.8)$$

where $\mu_t = \mathcal{L}(X_t^{s, \mu} | \mathcal{F}_t^0)$, $t \in [s, T]$. In addition, the derivatives $D_x^k X_t^{s, x, \mu}$ exist for $k = 2, 3, 4$. Furthermore, for every $p \geq 2$, there exists a constant $C_p > 0$ such that for all $t \in [s, T]$, $x, x' \in \mathbb{R}$ and $\eta, \eta' \in L^2(\mathcal{F}_s^1)$ with $\mathcal{L}(\eta) = \mu$ and $\mathcal{L}(\eta') = \mu'$

$$\mathbb{E} \left[\sup_{t \in [s, T]} |D_x^k X_t^{s, x, \mu}|^p \right] \leq C_p \quad (2.7.9)$$

$$\mathbb{E} \left[\sup_{t \in [s, T]} |D_x^k X_t^{s, x, \mu} - D_x^k X_t^{s, x', \mu'}|^p \right] \leq C_p (|x - x'|^p + \mathbf{d}_2(\mu, \mu')^p) \quad (2.7.10)$$

for $k = 1, 2, 3, 4$.

Proof. Note that (2.7.8) is obtained by formally differentiating (2.7.2). Let $D_x X_t^{s, x, \mu}$ be its solution. Then, by a standard Grönwall inequality argument (after invoking the BDG inequality the regularity of B , Σ and Σ^0 and (2.7.4), (2.7.5)) we get that

$$\mathbb{E} \left[\sup_{t \in [s, T]} |X_t^{s, x+h, \mu} - X_t^{s, x, \mu} - h D_x X_t^{s, x, \mu}|^2 \right] = o(|h|^2), \quad (2.7.11)$$

therefore $X_t^{s, x, \mu}$ is differentiable with respect to x , in the L^2 -sense, and its derivative is equal to $D_x X_t^{s, x, \mu}$. Now, for $k = 1$, estimates (2.7.9) and (2.7.23) follow from standard estimates for SDEs by (2.7.8) due to the regularity of B , Σ and Σ^0 .

For $k = 2, 3, 4$ the existence of $D_x^k X_t^{s, x, \mu}$ follows in a similar way by formally differentiating (2.7.8) ($l = 1, 2, 3$ times, respectively) and then using the regularity of B , Σ and Σ^0 deriving an estimate like (2.7.11). For $k = 2, 3, 4$, estimates (2.7.9) and (2.7.10) follow by standard estimates for the SDEs satisfied by $D_x^k X_t^{s, x, \mu}$ and estimates (2.7.9) and (2.7.10) for the previous derivatives. $\square$

Next we study the regularity of the map $\eta \mapsto X_t^{s, x, \mu}$, where $\mathcal{L}(\eta) = \mu$. In particular, we show that this map is Fréchet differentiable with Fréchet derivative given by $\eta \mapsto \left(\xi \mapsto \widetilde{\mathbb{E}}^1[U_t^{s, x, \mu}(\widetilde{\eta})\widetilde{\xi}] \right)$, where $U_t^{s, x, \mu}(y)$ is adapted to $(\mathcal{F}_t)_{t \in [s, T]}$ and where $(\widetilde{\eta}, \widetilde{\xi})$ is a copy of (η, ξ) defined over the probability space $(\widetilde{\Omega}^1, \widetilde{\mathcal{F}}^1, \widetilde{\mathbb{P}}^1)$ (recall Notation (ii) before Theorem 2.7.1). By Definition 2.3.1 the L -derivative $D_\mu X_t^{s, x, \mu}$ is $D_\mu X_t^{s, x, \mu}(y) := U_t^{s, x, \mu}(y)$, $t \in [s, T]$, $x, y \in \mathbb{R}$.

Proposition 2.7.3. *Let $y \in \mathbb{R}$, $(s, x, \mu) \in [0, T] \times \mathbb{R} \times \mathcal{P}_2(\mathbb{R})$, $X_t^{s, \mu}$ and $X_t^{s, x, \mu}$ be as in (2.7.1), (2.7.2) and $U_t^{s, x, \mu}(y)$, $U_t^{s, \mu}(y)$ be the unique solution to the system of (uncoupled)*

SDEs

$$\begin{aligned}
U_t^{s,x,\mu}(y) &= \int_s^t D_x B(X_r^{s,x,\mu}, \mu_r) U_r^{s,x,\mu}(y) dr \\
&+ \int_s^t \tilde{\mathbb{E}}^1 [D_\mu B(X_r^{s,x,\mu}, \mu_r, \tilde{X}_r^{s,y,\mu}) \cdot D_x \tilde{X}_r^{s,y,\mu} + D_\mu B(X_r^{s,x,\mu}, \mu_r, \tilde{X}_r^{s,\mu}) \cdot \tilde{U}_r^{s,\mathcal{L}(\tilde{\eta})}(y)] dr \\
&+ \int_s^t \tilde{\mathbb{E}}^1 [D_\mu \Sigma^1(X_r^{s,x,\mu}, \mu_r, \tilde{X}_r^{s,y,\mu}) \cdot D_x \tilde{X}_r^{s,y,\mu} + D_\mu \Sigma^1(X_r^{s,x,\mu}, \mu_r, \tilde{X}_r^{s,\mu}) \cdot \tilde{U}_r^{s,\mathcal{L}(\tilde{\eta})}(y)] dW_r \\
&+ \int_s^t \tilde{\mathbb{E}}^1 [D_\mu \Sigma^0(X_r^{s,x,\mu}, \mu_r, \tilde{X}_r^{s,y,\mu}) \cdot D_x \tilde{X}_r^{s,y,\mu} + D_\mu \Sigma^0(X_r^{s,x,\mu}, \mu_r, \tilde{X}_r^{s,\mu}) \cdot \tilde{U}_r^{s,\mathcal{L}(\tilde{\eta})}(y)] dB_r^0 \\
&+ \int_s^t D_x \Sigma^1(X_r^{s,x,\mu}, \mu_r) U_r^{s,x,\mu}(y) dW_r + \int_s^t D_x \Sigma^0(X_r^{s,x,\mu}, \mu_r) U_r^{s,x,\mu}(y) dB_r^0, \tag{2.7.12}
\end{aligned}$$

$$\begin{aligned}
U_t^{s,\mu}(y) &= \int_s^t D_x B(X_r^{s,\mu}, \mu_r) U_r^{s,\mu}(y) dr \\
&+ \int_s^t \tilde{\mathbb{E}}^1 [D_\mu B(X_r^{s,\mu}, \mu_r, \tilde{X}_r^{s,y,\mu}) \cdot D_x \tilde{X}_r^{s,y,\mu} + D_\mu B(X_r^{s,\mu}, \mu_r, \tilde{X}_r^{s,\mu}) \cdot \tilde{U}_r^{s,\mathcal{L}(\tilde{\eta})}(y)] dr \\
&+ \int_s^t \tilde{\mathbb{E}}^1 [D_\mu \Sigma^1(X_r^{s,\mu}, \mu_r, \tilde{X}_r^{s,y,\mu}) \cdot D_x \tilde{X}_r^{s,y,\mu} + D_\mu \Sigma^1(X_r^{s,\mu}, \mu_r, \tilde{X}_r^{s,\mu}) \cdot \tilde{U}_r^{s,\mathcal{L}(\tilde{\eta})}(y)] dW_r \\
&+ \int_s^t \tilde{\mathbb{E}}^1 [D_\mu \Sigma^0(X_r^{s,\mu}, \mu_r, \tilde{X}_r^{s,y,\mu}) \cdot D_x \tilde{X}_r^{s,y,\mu} + D_\mu \Sigma^0(X_r^{s,\mu}, \mu_r, \tilde{X}_r^{s,\mu}) \cdot \tilde{U}_r^{s,\mathcal{L}(\tilde{\eta})}(y)] dB_r^0 \\
&+ \int_s^t D_x \Sigma^1(X_r^{s,\mu}, \mu_r) U_r^{s,\mu}(y) dW_r + \int_s^t D_x \Sigma^0(X_r^{s,\mu}, \mu_r) U_r^{s,\mu}(y) dB_r^0, \tag{2.7.13}
\end{aligned}$$

where $\eta \in L^2(\mathcal{F}_s^1)$ with $\mathcal{L}(\eta) = \mu$, $\mu_t = \mathcal{L}(X_t^{s,\mu} | \mathcal{F}_t^0)$, and Notation (i), (ii) (before Theorem 2.7.1) are used. Then, the Wasserstein derivative $D_\mu X_t^{s,x,\mu}(y)$, $y \in \mathbb{R}$, exists and is equal to $U_t^{s,x,\mu}(y)$. Moreover, for all $p \geq 2$, there exists a constant $C_p > 0$ such that for all $x, x', y, y' \in \mathbb{R}$ and $\eta, \eta' \in L^2(\mathcal{F}_s)$ with $\mathcal{L}(\eta) = \mu$, $\mathcal{L}(\eta') = \mu'$ the following estimates hold

$$\mathbb{E} \left[\sup_{t \in [s, T]} |D_\mu X_t^{s,x,\mu}(y)|^p \right] \leq C_p, \tag{2.7.14}$$

$$\begin{aligned}
&\mathbb{E} \left[\sup_{t \in [s, T]} (|D_\mu X_t^{s,x,\mu}(y) - D_\mu X_t^{s,x',\mu'}(y')|^p) \right] \\
&\leq C_p (|x - x'|^p + |y - y'|^p + \mathbf{d}_2(\mu, \mu')^p). \tag{2.7.15}
\end{aligned}$$

Proof. In the proof, for simplicity, we assume that $B = 0$ and that Σ^1, Σ^0 depend only on the measure argument. The extension to the more general case stated above follows easily by doing the analogous calculations.

Let $\eta, \xi \in L^2(\mathcal{F}_s^1)$ and $\tilde{\eta}, \tilde{\xi}$ be copies of η, ξ defined on $(\tilde{\Omega}^1, \tilde{\mathcal{F}}^1, \tilde{\mathbb{P}}^1)$.

Step 1: We first show that for all $(t, x) \in [s, T] \times \mathbb{R}$ the L^2 - Gâteaux derivative

$$\lim_{h \rightarrow 0} \frac{1}{h} (X_t^{s,x,\mathcal{L}(\eta+h\xi)} - X_t^{s,x,\mathcal{L}(\eta)})$$

exists. In fact, we will show that if $Z_t^{s,x,\mathcal{L}(\eta)}(\xi)$ is the solution of the SDE

$$\begin{aligned} & Z_t^{s,x,\mathcal{L}(\eta)}(\xi) \\ &= \int_s^t \mathbb{E}^1[D_\mu \Sigma^1(\mathcal{L}(X_r^{s,\mathcal{L}(\eta)} | \mathcal{F}_r^0), \tilde{X}_r^{s,\mathcal{L}(\tilde{\eta})})(D_x X_r^{s,y,\mathcal{L}(\tilde{\eta})} \tilde{\xi} + \tilde{Z}_r^{s,y,\mathcal{L}(\tilde{\eta})}(\tilde{\xi})) \Big|_{y=\tilde{\eta}}] dW_r \\ &+ \int_s^t \mathbb{E}^1[D_\mu \Sigma^0(\mathcal{L}(X_r^{s,\mathcal{L}(\eta)} | \mathcal{F}_r^0), \tilde{X}_r^{s,\mathcal{L}(\tilde{\eta})})(D_x X_r^{s,y,\mathcal{L}(\tilde{\eta})} \tilde{\xi} + \tilde{Z}_r^{s,y,\mathcal{L}(\tilde{\eta})}(\tilde{\xi})) \Big|_{y=\tilde{\eta}}] dB_r^0, \end{aligned} \quad (2.7.16)$$

which is the linearization of (2.7.2), then

$$\mathbb{E} \left[\sup_{t \in [s, T]} |X_t^{s,x,\mathcal{L}(\eta+h\xi)} - X_t^{s,x,\mathcal{L}(\eta)} - hZ_t^{s,x,\mathcal{L}(\eta)}(\xi)|^2 \right] \leq Ch^4 \mathbb{E}[\xi^2]^2 \quad (2.7.17)$$

for some constant $C > 0$ and any $h \in \mathbb{R}$. Indeed, due to estimates (2.7.3), (2.7.4) and (2.7.5), following the calculation from Lemma 4.2 of [BDL17] adjusted to accommodate the terms involving the common noise, we have

$$\begin{aligned} & X_t^{s,x,\mathcal{L}(\eta+h\xi)} - X_t^{s,x,\mathcal{L}(\eta)} - hZ_t^{s,x,\mathcal{L}(\eta)}(\xi) = R(t, x, h) \\ &+ \int_s^t \mathbb{E}^1[D_\mu \Sigma^1(\mathcal{L}(X_r^{s,\mathcal{L}(\eta)} | \mathcal{F}_r^0), \tilde{X}_r^{s,\mathcal{L}(\tilde{\eta})}) \\ &\quad \cdot (\tilde{X}_r^{s,\mathcal{L}(\tilde{\eta}+h\tilde{\xi})} - \tilde{X}_r^{s,\mathcal{L}(\tilde{\eta})} - h(D_x \tilde{X}_r^{s,z,\mathcal{L}(\tilde{\eta})} \tilde{\xi} + \tilde{Z}_r^{s,z,\mathcal{L}(\tilde{\eta})}(\tilde{\xi})) \Big|_{z=\tilde{\eta}}] dW_r \end{aligned}$$

$$\begin{aligned}
& + \int_s^t \tilde{\mathbb{E}}^1 [D_\mu \Sigma^0(\mathcal{L}(X_r^{s,\mathcal{L}(\eta)} | \mathcal{F}_r^0), \tilde{X}_r^{s,\mathcal{L}(\tilde{\eta})}) \\
& \cdot (\tilde{X}_r^{s,\mathcal{L}(\tilde{\eta}+h\tilde{\xi})} - \tilde{X}_r^{t,\mathcal{L}(\tilde{\eta})} - h(D_x \tilde{X}_r^{t,z,\mathcal{L}(\eta)} \tilde{\xi} + \tilde{Z}_r^{t,z,\mathcal{L}(\tilde{\eta})}(\tilde{\xi})) \Big|_{z=\tilde{\eta}}] dB_r^0, \tag{2.7.18}
\end{aligned}$$

where $R(t, x, h)$ satisfies

$$\mathbb{E} \left[\sup_{t \in [s, T]} |R(t, x, h)|^2 \right] \leq Ch^4 \mathbb{E}[|\xi|^2]^2.$$

Therefore, to prove (2.7.17), due to the boundness of $D_\mu \Sigma^1$ and $D_\mu \Sigma^0$ we only need to bound the expectation of the absolute of the quantity

$$\tilde{X}_r^{s,\mathcal{L}(\tilde{\eta}+h\tilde{\xi})} - \tilde{X}_r^{s,\mathcal{L}(\tilde{\eta})} - h(D_x \tilde{X}_r^{s,z,\mathcal{L}(\eta)} \tilde{\xi} + \tilde{Z}_r^{s,z,\mathcal{L}(\tilde{\eta})}(\tilde{\xi})) \Big|_{z=\tilde{\eta}} := I_1(r, h) + I_2(r, h),$$

which appears inside the integrals in (2.7.18). Here

$$\begin{aligned}
I_1(r, h) &= \tilde{X}_r^{s,z,\mathcal{L}(\tilde{\eta}+h\tilde{\xi})} - \tilde{X}_r^{s,\mathcal{L}(\tilde{\eta})} - h\tilde{Z}_r^{s,z,\mathcal{L}(\tilde{\eta})}(\tilde{\xi}), \\
I_2(r, h) &= \tilde{X}_r^{s,\tilde{\eta}+h\tilde{\xi},\mathcal{L}(\eta+h\xi)} - \tilde{X}_r^{s,\tilde{\eta},\mathcal{L}(\eta+h\xi)} - h\tilde{\xi} D_x \tilde{X}_r^{s,\tilde{\eta},\mathcal{L}(\eta)}.
\end{aligned}$$

We also define

$$A(r) := \sup_{x \in \mathbb{R}} \tilde{\mathbb{E}}[|\tilde{X}_r^{s,x,\mathcal{L}(\tilde{\eta}+h\tilde{\xi})} - \tilde{X}_r^{s,x,\mathcal{L}(\tilde{\eta})} - h\tilde{Z}_r^{s,x,\mathcal{L}(\tilde{\eta})}(\tilde{\xi})|^2].$$

We have

$$\begin{aligned}
\tilde{\mathbb{E}}[|I_1(r, h)|]^2 &= \tilde{\mathbb{E}}[|\tilde{X}_r^{s,z,\mathcal{L}(\tilde{\eta}+h\tilde{\xi})} - \tilde{X}_r^{s,\mathcal{L}(\tilde{\eta})} - h\tilde{Z}_r^{s,z,\mathcal{L}(\tilde{\eta})}(\tilde{\xi})|^2] \Big|_{z=\tilde{\eta}} \\
&\leq \sup_{x \in \mathbb{R}} \tilde{\mathbb{E}}[|\tilde{X}_r^{s,x,\mathcal{L}(\tilde{\eta}+h\tilde{\xi})} - \tilde{X}_r^{s,x,\mathcal{L}(\tilde{\eta})} - h\tilde{Z}_r^{s,x,\mathcal{L}(\tilde{\eta})}(\tilde{\xi})|^2] \quad \text{and} \\
\tilde{\mathbb{E}}[|I_2(r, h)|]^2 &= \tilde{\mathbb{E}}[|\tilde{X}_r^{s,\tilde{\eta}+h\tilde{\xi},\mathcal{L}(\eta+h\xi)} - \tilde{X}_r^{s,\tilde{\eta},\mathcal{L}(\eta+h\xi)} - h\tilde{\xi} D_x \tilde{X}_r^{s,\tilde{\eta},\mathcal{L}(\eta)}|^2]
\end{aligned}$$

$$\leq Ch^4 \mathbb{E}[|\xi|^2]^2,$$

where in the last inequality we used the fundamental theorem of calculus for the first two terms and (2.7.10). We combine these inequalities with (2.7.18) and we get

$$A(t) \leq Ch^4 \mathbb{E}[|\eta|^2]^2 + C \int_s^t A(r) dr.$$

(2.7.17) follows from Grönwall's inequality.

Step 2: Let $\widehat{\eta}, \widehat{\xi}$ be independent copies of η, ξ and $\widetilde{\eta}, \widetilde{\xi}$ defined on a new probability space $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{\mathbb{P}})$, which is a copy of $(\Omega^1, \mathcal{F}^1, \mathbb{P}^1)$. We claim that

$$Z_t^{s,x,\mathcal{L}(\eta)}(\xi) = \widehat{\mathbb{E}}[U_t^{s,x,\mu}(\widehat{\eta})\widehat{\xi}], \text{ a.s.}$$

The proof is a standard calculation involving Fubini's theorem. Indeed, we can easily see that $Z_t^{s,\mathcal{L}(\eta)}(\xi) := Z_t^{s,x,\mathcal{L}(\eta)}(\xi)|_{x=\eta}$ and $\widehat{\mathbb{E}}[U_t^{s,\mathcal{L}(\eta)}(\widehat{\eta})\widehat{\xi}]$ satisfy the same SDE ((2.7.16) for $x = \eta$) and hence, by uniqueness, we must have $Z_t^{s,\mathcal{L}(\eta)}(\xi) = \widehat{\mathbb{E}}[U_t^{s,\mathcal{L}(\eta)}(\widehat{\eta})\widehat{\xi}]$, a.s. Using this, we can verify that $Z_t^{s,x,\mathcal{L}(\eta)}(\xi)$ satisfies SDE as $\widehat{\mathbb{E}}[U_t^{s,x,\mu}(\widehat{\eta})\widehat{\xi}]$. The conclusion follows.

Step 3: Let $p \geq 2$, $x, x', y, y' \in \mathbb{R}$ and $\eta, \eta' \in L^2(\mathcal{F}_s^1)$ with $\mathcal{L}(\eta) = \mu$, $\mathcal{L}(\eta') = \mu'$. By combining (2.7.9), (2.7.10) and (2.7.5) with standard arguments involving Grönwall's and BDG inequalities, we deduce that there exists a constant $C_p > 0$ such that

$$\mathbb{E} \left[\sup_{t \in [s, T]} |U_t^{s,x,\mu}(y)|^p \right] \leq C_p, \quad (2.7.19)$$

$$\mathbb{E} \left[\sup_{t \in [s, T]} (|U_t^{s,x,\eta}(y) - U_t^{s,x',\eta'}(y')|^p) \right] \leq C_p (|x - x'|^p + |y - y'|^p + \mathbf{d}_2(\mu, \mu')^p). \quad (2.7.20)$$

We can also easily prove that SDE (2.7.13) and estimates (2.7.3), (2.7.5) and (2.7.9), (2.7.10)

imply

$$\mathbb{E} \left[\sup_{t \in [s, T]} |U_t^{s, \mu}(y) - U_t^{s, \mu'}(y')|^p \right] \leq C_p(|y - y'|^p + \mathbf{d}_2(\mu, \mu')^p) \quad (2.7.21)$$

By the relation proved in step 2 and (2.7.20), we can easily deduce that the Gâteaux derivative $Z_t^{s, x, \mathcal{L}(\eta)}$ is continuous in η , therefore it is the Fréchet derivative. Thus, the map $\eta \mapsto X_t^{s, x, \eta} = X_t^{s, x, \mathcal{L}(\eta)}$ is Fréchet differentiable, so we can define the L -derivative of the map $\mu \mapsto X_t^{s, x, \mu}$. By step 2, we have $D_\mu X_t^{s, x, \mu}(y) = U_t^{s, x, \mu}(y)$. The proof is complete. $\square$

Remark 2.7.4. (i) From estimate (2.7.20), we see that $U^{s, x, \eta}$ depends on η only through $\mathcal{L}(\eta)$, hence we have the identification $U_t^{s, x, \mathcal{L}(\eta)} = U_t^{s, x, \eta}$ that we were using since the beginning throughout the proof. The same is true for $U_t^{s, \mathcal{L}(\eta)} = U_t^{s, \eta}$, due to (2.7.21).

(ii) Let $y \in \mathbb{R}$. Substituting $x = \eta$ in (2.7.12), we deduce from the uniqueness of the solution that $U_t^{s, \mathcal{L}(\eta)}(y) = U_t^{s, x, \mathcal{L}(\eta)}(y) \Big|_{x=\eta}$, $\mathbb{P}$ -a.s. Using the same argument, we can also substitute $\mathcal{F}_s^1$ -measurable random variables for y in $U_t^{s, \mathcal{L}(\eta)}(y)$.

Corollary 2.7.5. *Let $X_t^{s, \mu}$ be as in (2.7.1). Then, for any $l \in \{0, 1, 2, 3, 4\}$ the map $\mu \mapsto D_x^l X_t^{s, x, \mu}$ satisfies the property that for any multi-index (n, β) such that $k = |(n, \beta)| \leq 4 - l$, $D^{(n, \beta)} D_x^l X_t^{s, x, \mu}$ exists. In addition, if $p \geq 2$ this derivative satisfies*

$$\mathbb{E} \left[\sup_{t \in [s, T]} |D^{(n, \beta)} D_x^l X_t^{s, x, \mu}(z_1, \dots, z_n)|^p \right] \leq C_p, \quad (2.7.22)$$

$$\begin{aligned} & \mathbb{E} \left[\sup_{t \in [s, T]} |D^{(n, \beta)} D_x^l X_t^{s, x, \mu}(z_1, \dots, z_n) - D^{(n, \beta)} D_x^l X_t^{s, x', \mu'}(z'_1, \dots, z'_n)|^p \right] \\ & \leq C_p(|x - x'|^p + \sum_{i=1}^n |z_i - z'_i|^p + \mathbf{d}_2(\mu, \mu')^p) \end{aligned} \quad (2.7.23)$$

for any $x, x', z_1, z'_1, \dots, z_n, z'_n \in \mathbb{R}$ and $\mu, \mu' \in \mathcal{P}_2(\mathbb{R})$ for some constant $C_p > 0$.

Proof. For $l = 0$ and $k = 0$ the result follows by relations (2.7.3), (2.7.5). For $l = 0$ and $k = 1$ the result follows by Proposition 2.7.3. For $l = 0$ and $k \in \{2, 3, 4\}$ the result follows

by repeating the linearization argument from Proposition 2.7.3 to the SDEs satisfied by the L -derivatives $D_\mu^i X_t^{s,x,\mu}(y)$, $i = 1, 2, \dots, k-1$ and by induction. For $l \in \{1, 2, 3, 4\}$ the result follows from the same argument as in Proposition 2.7.2 applied to $D_x^l X_t^{s,x,\mu}$ and estimates (2.7.9), (2.7.10). $\square$

We are now ready to prove Theorem 2.7.1.

Proof. Let $t \in [0, T]$, $\eta \in L^2(\mathcal{F}_t^1)$ with $\mathcal{L}(\eta) = \mu$ and $\xi \in L^2(\mathcal{F}_t^1)$.

Step 1: Our first step is to show that for any multi-index (n, β) such that $|(n, \beta)| \leq 4$ the derivative $D^{(n, \beta)} v(t, \mu, y_1, \dots, y_n)$ exists and it is bounded by $C \|G\|_{\mathcal{M}_b^\xi}$ for some constant C (independent of (n, β) and Lipschitz (with respect to μ and $y_1, \dots, y_n$). This will give us $v(t, \cdot) \in \mathcal{M}_{b,L}^\xi$ and the estimate on $\|v\|_{\mathcal{M}_b^\xi}$.

We observe that for $h \in \mathbb{R}$ and $s \in [t, T]$, due to Propositions 2.7.2 and 2.7.3, we have

$$\begin{aligned} D_h X_s^{t, \mathcal{L}(\eta+h\xi)} &= D_h \left(X_s^{t, x, \mathcal{L}(\eta+h\xi)} \Big|_{x=\eta+h\xi} \right) \\ &= \left(D_x X_s^{t, x, \mathcal{L}(\eta+h\xi)} \Big|_{x=\eta+h\xi} \right) \xi + \left(\lim_{r \rightarrow 0} \frac{X_s^{t, x, \mathcal{L}(\eta+(h+r)\xi)} - X_s^{t, x, \mathcal{L}(\eta+h\xi)}}{r} \right) \Big|_{x=\eta+h\xi}. \end{aligned}$$

Thus, the map $\eta \mapsto X_s^{t, \mathcal{L}(\eta)}$ is Fréchet differentiable with directional derivative

$$D_\eta (X_s^{t, \mathcal{L}(\eta)})(\xi) = D_h X_s^{t, \mathcal{L}(\eta+h\xi)} \Big|_{h=0} = \left(D_x X_s^{t, x, \mathcal{L}(\eta)} \Big|_{x=\eta} \right) \xi + D_\eta (X_s^{t, x, \mathcal{L}(\eta)})(\xi) \Big|_{x=\eta}.$$

We define the lifted functions $G' : L^2(\mathcal{F}_T^1) \rightarrow \mathbb{R}$ and $v'(t, \cdot) : L^2(\mathcal{F}_t^1) \rightarrow \mathbb{R}$ of G , $v(t, \cdot)$, respectively, which are given by $G'(\eta) = G(\mathcal{L}(\eta))$ and $v'(t, \eta) = v(t, \mathcal{L}(\eta))$. Therefore, by the chain rule of Fréchet differentiation we have

$$D_\eta v(t, \eta)(\xi) = \mathbb{E}^0 \left[DG'(X_T^{t, \mathcal{L}(\eta)})(D_\eta X_T^{t, \mathcal{L}(\eta)}(\xi)) \right]$$

$$\begin{aligned}
&= \mathbb{E}^0 \left[\mathbb{E}^1 \left[D_\mu G(\mathcal{L}(X_T^{t,\mathcal{L}(\eta)} | \mathcal{F}_T^0), X_T^{t,\mathcal{L}(\eta)}) \right. \right. \\
&\quad \left. \left. \left(\left(D_x X_s^{t,x,\mathcal{L}(\eta)} \right) \Big|_{x=\eta} \right) \xi + D_\eta (X_s^{t,x,\mathcal{L}(\eta)})(\xi) \Big|_{x=\eta} \right) \right] \right] \\
&= \mathbb{E}^0 \left[\mathbb{E}^1 \left[D_\mu G(\mathcal{L}(X_T^{t,\mathcal{L}(\eta)} | \mathcal{F}_T^0), X_T^{t,\mathcal{L}(\eta)}) \right. \right. \\
&\quad \left. \left. \left(\left(D_x X_s^{t,x,\mathcal{L}(\eta)} \right) \Big|_{x=\eta} \right) \xi + \widetilde{\mathbb{E}}^1 [D_\mu X_T^{t,x,\mathcal{L}(\eta)}(\tilde{\eta}) \Big|_{x=\eta} \tilde{\xi}] \right) \right] \right] \\
&= \widetilde{\mathbb{E}}^1 \left[\mathbb{E}^0 \left[\mathbb{E}^1 \left[D_\mu G(\mathcal{L}(X_T^{t,\mathcal{L}(\eta)} | \mathcal{F}_T^0), X_T^{t,x,\mathcal{L}(\eta)}) D_x X_s^{t,x,\mathcal{L}(\eta)} \Big|_{x=\tilde{\eta}} \tilde{\xi} \right] \right] \right] \\
&\quad + \widetilde{\mathbb{E}}^1 \left[\mathbb{E}^0 \left[\mathbb{E}^1 \left[D_\mu G(\mathcal{L}(X_T^{t,\mathcal{L}(\eta)} | \mathcal{F}_T^0), X_T^{t,\mathcal{L}(\eta)}) D_\mu X_T^{t,x,\mathcal{L}(\eta)}(\tilde{\eta}) \Big|_{x=\eta} \right] \right] \tilde{\xi} \right],
\end{aligned}$$

where in the last equality we used that η, ξ are $\mathcal{F}_t^1$ -measurable and that $\tilde{\eta}, \tilde{\xi}$ are their copies on $(\widetilde{\Omega}^1, \widetilde{\mathcal{F}}_t^1, \widetilde{\mathbb{P}}^1)$. Since $\mathcal{L}(\eta) = \mu$, this equality implies

$$\begin{aligned}
D_\mu v(t, \mu, y) &= \mathbb{E}^0 \left[\mathbb{E}^1 [D_\mu G(\mathcal{L}(X_T^{t,\mu} | \mathcal{F}_T^0), X_T^{t,y,\mu}) D_x X_T^{t,y,\mu} \right. \\
&\quad \left. + D_\mu G(\mathcal{L}(X_T^{t,\mu} | \mathcal{F}_T^0), X_T^{t,\mu}) U_T^{t,\mu}(y)] \right], \tag{2.7.24}
\end{aligned}$$

where $U_T^{t,\mu}$ is defined in Proposition 2.7.3. Since $G \in \mathcal{M}_b^\dot{\xi}$ and (2.7.9), (2.7.10), (2.7.14) and (2.7.21) hold, we see that $D_\mu v(t, \mu, y)$ is bounded by $C \|G\|_{\mathcal{M}_{b,L}^\dot{\xi}}$ (for some constant C) and is also Lipschitz with respect to μ and y .

To verify the differentiability of higher order we may use Corollary 2.7.5. Indeed, to compute the higher order derivatives with respect to the measure variable, we lift $D_\mu G(\cdot, y)$, $D_\mu v(t, \cdot, y)$ to functions $(D_\mu G)'(\cdot, y)$, $(D_\mu v)'(t, \cdot, y)$ defined over $L^2(\mathcal{F}_t^1)$ so that

$$\begin{aligned}
(D_\mu v)'(t, \eta, y) &= \mathbb{E}^0 \left[\mathbb{E}^1 [(D_\mu G)'(X_T^{t,\mathcal{L}(\eta)}, X_T^{t,y,\mathcal{L}(\eta)}) D_x X_T^{t,y,\mathcal{L}(\eta)} + \right. \\
&\quad \left. (D_\mu G)'(X_T^{t,\mathcal{L}(\eta)}, X_T^{t,\mathcal{L}(\eta)}) D_\mu X_T^{t,x,\mathcal{L}(\eta)}(y) \Big|_{x=\eta}] \right],
\end{aligned}$$

and, as previously, we compute the Fréchet derivative, which will then give us $D_{\mu\mu}^2 v$. To com-

pute $D_{\mu\mu\mu}^3 v$ and $D_{\mu\mu\mu\mu}^4 v$, we follow the same procedure. The computation of the derivatives with respect to the rest of the variables is straightforward by Corollary 2.7.5. The fact that the higher order derivatives will be bounded and Lipschitz follows from (2.7.22) and (2.7.23).

Step 2: Our second step is to prove that $v(t, \mu)$ is differentiable with respect to t and that the derivative is bounded. Indeed, by the uniqueness of the conditional distribution we have

$$v(t, \mu) = \mathbb{E}^0 \left[G(\mathcal{L}(X_{T-t}^\mu | \mathcal{F}_{T-t}^0)) \right].$$

The desired differentiability now follows from Itô's formula (Theorem 4.14 of [CD⁺18c]), which provides us with a formula for $\dot{v}$ (equality (2.7.6)). The estimate

$$\sup_{t \in [0, T], \mu \in \mathcal{P}_2(\mathbb{R})} |\dot{v}(t, \mu)| \leq C \|G\|_{\mathcal{M}_b^\xi}$$

follows from that formula, the boundness of B , Σ^1 , Σ^0 and the boundness of G and its derivatives.

Step 3: We now show that v is the unique classical solution to (2.7.6). Indeed, let $\mu \in \mathcal{P}_2(\mathbb{R})$, $t \in [0, T]$, V satisfying (2.7.6) in the classical sense and $\mu_s = \mathcal{L}(X_s^{t, \mu} | \mathcal{F}_s^0)$. Then, by Itô's formula we have

$$\begin{aligned} \mathbb{E}^0 \left[V(T, \mu_T) - V(t, \mu_t) \right] &= \int_t^T \mathbb{E}^0 \left[-\dot{V}(s, \mu_s) - \int_{\mathbb{R}} B(x, \mu_s) \cdot D_\mu V(s, \mu_s, x) \mu_s(dx) \right] ds \\ &\quad + \int_t^T \mathbb{E}^0 \left[\frac{1}{2} (\Sigma^1 \Sigma^{1\top}(x, \mu) + \Sigma^0 \Sigma^{0\top}(x, \mu_s)) : D_x D_\mu V(s, \mu_s, x) \mu_s(dx) \right] ds \\ &\quad - \int_t^T \mathbb{E}^0 \left[\frac{1}{2} \int_{\mathbb{R}^2} \Sigma^0(x, \mu_s) \Sigma^{0\top}(x', \mu_s) : D_{\mu\mu}^2 v(t, \mu_s, x, x') \mu_s(dx) \mu_s(dx') \right] ds. \end{aligned}$$

Since V satisfies (2.7.6), the right hand side is 0, so

$$V(t, \mu) = \mathbb{E}^0[G(\mu_T)] = \mathbb{E}^0[G(\mathcal{L}(X_T^{t, \mu} | \mathcal{F}_T^0))] = v(t, \mu).$$

Step 4: We finally show estimate (2.7.7). Since $v \in \mathcal{M}_{b, L}^{\dot{\zeta}}$ and v satisfies (2.7.6) (step 3), we have that the derivatives $D_\mu \dot{v}(t, \mu, x)$, $D_x D_\mu \dot{v}(t, \mu, x)$ and $D_{\mu\mu}^2 \dot{v}(t, \mu, x_1, x_2)$ exist and can be computed by differentiating

$$\begin{aligned} & - \int_{\mathbb{R}^d} \left(B(x, \mu) \cdot D_\mu v(t, \mu, x) + \frac{1}{2} (\Sigma^1 \Sigma^{1\top} + \Sigma^0 \Sigma^{0\top})(x, \mu) : D_x D_\mu v(t, \mu, x) \right) \mu(dx) \\ & - \frac{1}{2} \int_{\mathbb{R}^{2d}} \Sigma^0(x, \mu) \Sigma^{0\top}(x', \mu) : D_{\mu\mu}^2 v(t, \mu, x, x') \mu(dx) \mu(dx'). \end{aligned}$$

Estimate (2.7.7) now follows from the boundness of B , Σ^1 , Σ^0 and their derivatives and the boundness of $\|v\|_{\mathcal{M}_b^{\dot{\zeta}}}$ derived in Step 1. $\square$

2.8 On the choice of the Assumptions

In this section we will provide some comments related to Remark 2.3.6 (vi).

We assume that in (2.1.1) instead of $\frac{\lambda}{\epsilon} f(X_t^\epsilon, Y_t^\epsilon, \mathcal{L}(X_t^\epsilon | \mathcal{F}_t^0))$, we have the coefficient $\frac{1}{\epsilon} b(X_t^\epsilon, Y_t^\epsilon, \mathcal{L}(X_t^\epsilon | \mathcal{F}_t^0))$ for some function $b \in \mathcal{M}_p^{\hat{\zeta}}$ such that $\bar{b}(x, \mu) = 0$. For simplicity, we also assume that $c = g = 0$. We consider the function $\Phi \in \mathcal{M}_p^{\hat{\zeta}}(\mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d); \mathbb{R}^d)$ which is such that $L_0^{x, \mu} \Phi(x, y, \mu) = -b(x, y, \mu)$. Note that such Φ exists by Proposition 2.4.1.

Keeping the same notation as in section 1.2, assuming that the convergence holds, we want to find a function u such that $u^\epsilon(t, m)$ converges to $u(t, \pi_*^1 m)$ as $\epsilon \rightarrow 0$. By repeating the construction of the correctors u_1 , u_2 , which, for reference, is sketched in section 5.1, we find that $u_1(t, m) = \int_{\mathbb{R}^{2d}} \Phi(x, y, \mu) \cdot D_\mu u(t, \pi_*^1 m, x) m(dx, dy)$ with $\mu = \pi_*^1 m$, and, to

construct u_2 ,

$$\begin{cases} \dot{u}(t, \mu) + \mathcal{L}u(t, \mu) = 0, & (t, \mu) \in [0, T) \times \mathcal{P}_2(\mathbb{R}^d), \\ u(T, \mu) = G(\mu), & \mu \in \mathcal{P}_2(\mathbb{R}^d), \end{cases} \quad (2.8.1)$$

with

$$\begin{aligned} \mathcal{L}u(t, \mu) = & \int_{\mathbb{R}^d} \left(\overline{(D_x \Phi b)} + \overline{(\tau_1 \sigma^\top : D_{xy}^2 \Phi)} \right) (x, \mu) \cdot D_\mu u(t, \mu, x) \mu(dx) \\ & + \int_{\mathbb{R}^d} \overline{(\sigma_1^0 \sigma_2^{0\top} : D_{xy}^2 \Phi)} (x, \mu) \cdot D_\mu u(t, \mu, x) \mu(dx) \\ & + \int_{\mathbb{R}^{2d}} \int_{\mathbb{R}^d} \overline{D_\mu \Phi(x, y, \mu, x') b(x', y', \mu)} \mu(dx') \cdot D_\mu u(t, \mu, x) \mu(dx) \\ & + \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \overline{\sigma_2^0(x, \mu) \sigma_1^{0\top}(x', \mu) : D_y D_\mu \Phi(x, y, \mu, x')} \mu(dx') \cdot D_\mu u(t, \mu, x) \mu(dx) \\ & + \frac{1}{2} \int_{\mathbb{R}^{2d}} \left(\overline{\sigma \sigma^\top} + \frac{\overline{b \otimes \Phi} + \overline{(b \otimes \Phi)^\top}}{2} \right) (x, \mu) : D_x D_\mu u(t, \mu, x) \mu(dx) \\ & + \frac{1}{2} \int_{\mathbb{R}^d} \frac{\overline{D_y \Phi \tau_1 \sigma^\top} + \overline{(D_y \Phi \tau_1 \sigma^\top)^\top}}{2} (x, \mu) : D_x D_\mu u(t, \mu, x) \mu(dx) \\ & + \frac{1}{2} \int_{\mathbb{R}^d} \left(\sigma_1^0 \sigma_1^{0\top} + 2\sigma_1^0 (\overline{D_y \Phi \sigma_2^0})^\top \right) (x, \mu) : D_x D_\mu u(t, \mu, x) \mu(dx) \\ & + \frac{1}{2} \int_{\mathbb{R}^{2d}} (\sigma_1^0(x, \mu) \sigma_1^{0\top}(x', \mu) + \overline{(D_y \Phi \sigma_2^0)(x, y, \mu) (D_y \Phi \sigma_2^0)^\top(x', y', \mu)}) : D_{\mu\mu}^2 u(t, \mu, x, x') \mu(dx) \mu(dx') \\ & + \int_{\mathbb{R}^{2d}} (\sigma_1^0(x, \mu) \overline{(D_y \Phi \sigma_2^0)^\top(x', y', \mu)}) : D_{\mu\mu}^2 u(t, \mu, x, x') \mu(dx) \mu(dx'), \end{aligned}$$

where we use the notation $\bar{h}(x, x') = \int_{\mathbb{R}^{2d}} h(x, x', y, y') \Pi_{x, x', \mu}(dy, dy')$ with $\Pi_{x, x', \mu}$ being the invariant measure defined after (2.4.1). We observe that u satisfies an equation of the form

$$\begin{aligned} \dot{u}(t, \mu) + \int_{\mathbb{R}^d} (B(x, \mu) \cdot D_\mu u(t, \mu, x) + \Sigma_1(x, \mu) : D_x D_\mu u(t, \mu, x)) \mu(dx) \\ + \frac{1}{2} \int_{\mathbb{R}^{2d}} \Sigma_0(x, x', \mu) : D_{\mu\mu}^2 u(t, \mu, x, x') \mu(dx) \mu(dx') = 0, \end{aligned} \quad (2.8.2)$$

with B , Σ_1 , Σ_0 being regular. To our knowledge, it is not known if, in general, there exists solution u to such a problem and if the solution will be regular as well, unless we can identify

(2.8.2) with a McKean-Vlasov control equation associated with a conditional McKean-Vlasov process. To do that, Σ_0 has to satisfy $\Sigma_0(x, x', \mu) = \Sigma_2(x, \mu)\Sigma_2^\top(x', \mu)$ for some other matrix valued function Σ_2 , for every $x, x' \in \mathbb{R}^d$ and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. In our case

$$\begin{aligned}\Sigma_0(x, x', \mu) &= \sigma_1^0(x, \mu)\sigma_1^{0\top}(x', \mu) + \overline{(D_y\Phi\sigma_2^0)(x, y, \mu)(D_y\Phi\sigma_2^0)^\top(x', y', \mu)} \\ &\quad + \sigma_1^0(x, \mu)\overline{(D_y\Phi\sigma_2^0)^\top(x', \mu)} + \overline{(D_y\Phi\sigma_2^0)(x, \mu)}\sigma_1^{0\top}(x', \mu).\end{aligned}$$

Even though this is close to being equal to

$$(\sigma_1^0(x, \mu) + \overline{D_y\Phi\sigma_2^0}(x, \mu))(\sigma_1^0(x', \mu) + \overline{D_y\Phi\sigma_2^0}(x', \mu))^\top,$$

this is not true because

$$\overline{(D_y\Phi\sigma_2^0)(x, y, \mu)(D_y\Phi\sigma_2^0)^\top(x', y', \mu)} \neq \overline{D_y\Phi\sigma_2^0}(x, \mu)(\overline{D_y\Phi\sigma_2^0})^\top(x', \mu), \quad (2.8.3)$$

since $\Pi_{x, x', \mu}$ is not the product measure $\pi_{x, \mu} \otimes \pi_{x', \mu}$. However, if $D_y\Phi$ is independent of y (which is the case in our result), since σ_2^0 is also independent of y , (2.8.3) holds as an equality, as we are averaging over a function which is independent of y, y' . This justifies our choice of coefficients in (2.1.1).

Acknowledgments.

During the course of this work the author was partially supported by Panagiotis E. Souganidis' National Science Foundation DMS-1266383 and DMS-1600129, Office for Naval Research N000141712095 and Air Force Office for Scientific Research FA9550-18-1-0494 grants. The author wishes to thank Panagiotis E. Souganidis, Ben Seeger and Nikiforos Mimikos-Stamatopoulos for valuable discussions and the Institute of Mathematical and Statistical Innovation (IMSI) for its hospitality.

CHAPTER 3

SINGULAR PERTURBATION OF HAMILTON-JACOBI-BELLMAN EQUATIONS IN THE WASSERSTEIN SPACE

3.1 Introduction

3.1.1 Background

Let $\varepsilon, T > 0$, $d \in \mathbb{N}^*$ and $\mathcal{F}, \mathcal{G} : \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$ be two functions, where $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$ is the d -dimensional torus and $\mathcal{P}(\mathbb{T}^d)$ is the space of probability measures over $\mathbb{T}^d$. We are interested in the asymptotic behavior as $\varepsilon \rightarrow 0$ of the solution¹ $U^\varepsilon : [0, T] \times \mathcal{P}(\mathbb{T}^{2d}) \rightarrow \mathbb{R}$ of the Hamilton-Jacobi equation posed over $[0, T] \times \mathcal{P}(\mathbb{T}^{2d})$:

$$\left\{ \begin{array}{l} -\partial_t U^\varepsilon + \int_{\mathbb{T}^{2d}} H^\varepsilon(x, y, D_m U^\varepsilon(t, m, x, y)) m(dx, dy) \\ - \int_{\mathbb{T}^{2d}} \text{tr} \left((\sigma^\varepsilon \sigma^{\varepsilon\top} + \sigma_0^\varepsilon \sigma_0^{\varepsilon\top}) D_{(x,y)} D_m U^\varepsilon(t, m, x, y) \right) m(dx, dy) \\ - \int_{\mathbb{T}^{2d}} \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_0^\varepsilon \sigma_0^{\varepsilon\top} D_{mm}^2 U^\varepsilon(t, m, x, y, x', y') \right) m(dx, dy) m(dx', dy') \\ = \mathcal{F}(\pi_*^1 m), \\ U^\varepsilon(T, m) = \mathcal{G}(\pi_*^1 m), \end{array} \right. \quad (3.1.1)$$

where $\pi_*^1 m \in \mathcal{P}(\mathbb{T}^d)$ is the the x -marginal of m and, for $x, y \in \mathbb{T}^d$, $\mathbf{p} = (p, q) \in \mathbb{R}^d \times \mathbb{R}^d$

$$H^\varepsilon(x, y, \mathbf{p}) = H^\varepsilon(x, y, p, q) := H\left(x, y, p, \frac{q}{\varepsilon}\right),$$

1. The precise definition is provided in the next section. $D_m U$ denotes the Wasserstein derivative; we refer to the previous chapter.

$$\sigma^\varepsilon = \begin{pmatrix} \sigma_1 & \mathbb{O}_{d \times d} \\ \mathbb{O}_{d \times d} & \frac{1}{\sqrt{\varepsilon}} \sigma_2 \end{pmatrix} = \begin{pmatrix} \mathbb{I}_{d \times d} & \mathbb{O}_{d \times d} \\ \mathbb{O}_{d \times d} & \frac{1}{\sqrt{\varepsilon}} \mathbb{I}_{d \times d} \end{pmatrix} \sigma, \quad \text{where } \sigma = \begin{pmatrix} \sigma_1 & \mathbb{O}_{d \times d} \\ \mathbb{O}_{d \times d} & \sigma_2 \end{pmatrix},$$

$$\text{and } \sigma_0^\varepsilon = \begin{pmatrix} \sigma_{0,1} & \mathbb{O}_{d \times d} \\ \frac{1}{\sqrt{\varepsilon}} \sigma_{0,2} & \mathbb{O}_{d \times d} \end{pmatrix},$$

with $H : \mathbb{T}^d \times \mathbb{T}^d \times \mathbb{R}^{2d} \rightarrow \mathbb{R}$ being a given Hamiltonian, $\sigma_1, \sigma_2, \sigma_{0,1}, \sigma_{0,2}$ given $d \times d$ matrices and $\mathbb{O}_{d \times d}, \mathbb{I}_{d \times d}$ the $d \times d$ zero matrix and identity matrix, respectively.

Hamilton-Jacobi equations in infinite dimensions, as (3.1.1), appear in mean field control theory and its twin, mean field game theory when H is convex in $\mathbf{p}$, and, also, in the study of zero-sum stochastic differential games [CP19], when H is not necessarily convex in $\mathbf{p}$. These theories, typically, study the asymptotic behavior of large populations of weakly interacting particles/agents. Each agent controls a state process which is impacted by an idiosyncratic Brownian noise and/or a common noise. In general, Hamilton-Jacobi equations in infinite dimensions do not have classical solutions even if the data is smooth (see [BC18] and [DS24, Example 2.4] for examples). As a consequence, it is essential to work with weaker notions of solutions such as monotone solutions or various notions of viscosity solutions, depending on the setup and the assumptions. For results in these directions, we refer to [Ber21, BLL24, CS21] for mean field games and [Ber23, GT19, CKT23, CGK⁺24, SY24, DS24, BEZ23, TTZ23, DJS23, CTQ25, BCE⁺25] for mean field control, among others.

In this paper, we consider viscosity solutions of (3.1.1) in the sense of test functions (Definition 3.3.5). In particular, it was proved in [DJS23] that there is a unique viscosity solution U^ε of (3.1.1). Our goal is to study the behavior of U^ε as $\varepsilon \rightarrow 0$. In the case where $\sigma_0^\varepsilon = \mathbb{O}_{2d \times 2d}$, the finite dimensional analogue of (3.1.1) is a Hamilton-Jacobi equation posed

in $\mathbb{T}^{2d}$ instead of $\mathcal{P}(\mathbb{T}^{2d})$ and takes the form

$$\begin{cases} -\partial_t u^\varepsilon + h\left(x, y, D_x v^\varepsilon, \frac{1}{\varepsilon} D_y u^\varepsilon\right) - \operatorname{tr}\left(\sigma_1 \sigma_1^\top D_{xx}^2 u^\varepsilon\right) - \frac{1}{\varepsilon} \operatorname{tr}\left(\sigma_2 \sigma_2^\top D_{yy}^2 u^\varepsilon\right) = 0, \\ u^\varepsilon(T, x, y) = g(x), \end{cases} \quad (3.1.2)$$

for a Hamiltonian h and some function $g : \mathbb{T}^d \rightarrow \mathbb{R}$. This problem was studied in [BB05] and, for fully nonlinear equations, in [AB03] and the references therein. The main observation is that the family of viscosity solutions of (3.1.2) $(u^\varepsilon)_{\varepsilon>0}$ is uniformly bounded, consequently we may consider the semi-limits

$$\begin{aligned} \bar{u}(t, x) &:= \limsup_{\varepsilon \rightarrow 0, (t', x') \rightarrow (t, x)} \sup_{y \in \mathbb{T}^d} u^\varepsilon(t', x', y), \text{ if } t < T, \\ \bar{u}(T, x) &:= \limsup_{(t', x') \rightarrow (T, x), t' < T} \bar{u}(t', x') \end{aligned}$$

and $\underline{u}$ (defined as $\bar{u}$ but with $\liminf$, $\inf$ instead $\limsup$, $\sup$). It is clear that $\bar{u} \geq \underline{u}$. On the other hand, the semi-limits $\bar{u}, \underline{u}$ can be proved to be a subsolution and a supersolution of the equation

$$\begin{cases} -\partial_t u + \bar{h}(x, D_x u) - \operatorname{tr}\left(\sigma_1 \sigma_1^\top D_{xx}^2 u\right) = 0, \\ u(T, x) = g(x), \end{cases} \quad (3.1.3)$$

respectively, where $\bar{h}$ is another function called the effective Hamiltonian. The effective Hamiltonian has the property that for $(x, p) \in \mathbb{T}^d \times \mathbb{R}^d$, the equation

$$-\operatorname{tr}\left(\sigma_2 \sigma_2^\top D_{yy}^2 v\right) + h(x, y, p, D_y v) = \bar{h}(x, p), \quad y \in \mathbb{T}^d$$

has a unique viscosity solution. By comparison principle, it is expected that $\bar{u} \leq \underline{u}$, hence $\bar{u} = \underline{u} := u$, which will be a viscosity solution of (3.1.3). In this paper we prove an analogous result for equations posed on the Wasserstein space $\mathcal{P}(\mathbb{T}^{2d})$. More specifically, we show the

following result for (3.1.1).

Theorem 3.1.1. *Assume (A) from section 3.2 is in place. Then, there exists a function $\bar{H} : \mathbb{T}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ such that for $(x, p) \in \mathbb{T}^d \times \mathbb{R}^d$, $\bar{H}(x, p)$ is the unique constant such that*

$$-tr\left((\sigma_2\sigma_2^\top + \sigma_{0,2}\sigma_{0,2}^\top)D_{yy}^2v\right) + H(x, y, p, D_yv) = \bar{H}(x, p), \quad y \in \mathbb{T}^d,$$

has a unique classical solution $v(y) = v(y; x, p)$ and the equation on $[0, T] \times \mathcal{P}(\mathbb{T}^d)$

$$\begin{cases} -\partial_t U + \int_{\mathbb{T}^d} \bar{H}(x, D_\mu U(t, \mu, x)) \mu(dx) \\ \quad - \int_{\mathbb{T}^d} tr\left((\sigma_1\sigma_1^\top + \sigma_{0,1}\sigma_{0,1}^\top)D_x D_\mu U(t, \mu, x)\right) \mu(dx) \\ \quad - \int_{\mathbb{T}^d} \int_{\mathbb{T}^d} tr\left(\sigma_{0,1}\sigma_{0,1}^\top D_{\mu\mu}^2 U(t, \mu, x, y)\right) \mu(dx) \mu(dy) = \mathcal{F}(\mu), \\ U(T, \mu) = \mathcal{G}(\mu) \end{cases} \quad (3.1.4)$$

has a unique viscosity solution $U : [0, T] \times \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$. Furthermore, $U^\varepsilon(t_0, m_0)$ converges as $\varepsilon \rightarrow 0$ to $U(t_0, \pi_1^* m_0)$ uniformly in $[0, T] \times \mathcal{P}(\mathbb{T}^{2d})$.

Remark 3.1.2. In the literature the function $\bar{H}(x, p)$ is called the effective Hamiltonian associated with the Hamiltonian $H(x, y, p, q)$. Therefore, $\bar{\mathcal{H}}(\mu, p) = \int_{\mathbb{T}^d} \bar{H}(x, p) \mu(dx)$, $(\mu, p) \in \mathcal{P}(\mathbb{T}^d) \times \mathbb{R}^d$ serves as the effective Hamiltonian associated to the Hamiltonian of (3.1.1).

3.1.2 Slow-fast systems and slow-fast stochastic optimal control problems

Apart from proving results from [AB03, BB05] in an infinite dimensional setting, the motivation to study this problem also comes from looking (3.1.1) through the lens of mean field control theory. We assume that A is a compact subset of $\mathbb{R}^{2d}$ and H is convex with

$$H(x, y, p, q) = \sup_{\alpha \in A} \{-p \cdot c(x, y, \alpha) - q \cdot f(x, y, \alpha) - L(x, y, \alpha)\} \quad (3.1.5)$$

for some functions $f, c : \mathbb{T}^d \times \mathbb{T}^d \times A \rightarrow \mathbb{R}^d$ and $L : \mathbb{T}^d \times \mathbb{T}^d \times A \rightarrow \mathbb{R}$ (assumption (A2)). It is known that $U^\varepsilon(t_0, m_0)$ can be characterized as the value function of an optimal control problem. To define the value function, we work on the canonical space $\Omega^0 := \mathbb{T}^{2d} \times \mathcal{C} \times \mathcal{C} \times \mathcal{C}$ with $\mathcal{C} = C([0, T]; \mathbb{R}^d)$. For $m \in \mathcal{P}(\mathbb{T}^{2d})$ we let $\mathbb{P}^m = m \otimes \mathbb{P}^0 \otimes \mathbb{P}^0 \otimes \mathbb{P}^0$, where $\mathbb{P}^0$ is the Wiener measure on $\mathcal{C}$. We write $(x, \omega_1, \omega_2, \omega^0)$ for the general element of Ω^0 and (ξ, B^1, B^2, B^0) the canonical random variable and processes on Ω^0 , i.e

$$\begin{aligned}\xi(x, \omega_1, \omega_2, \omega^0) &= x, \quad B_t^1(x, \omega_1, \omega_2, \omega^0) = \omega^1(t), \\ B_t^2(x, \omega_1, \omega_2, \omega^0) &= \omega^2(t), \quad B_t^0(x, \omega_1, \omega_2, \omega^0) = \omega^0(t).\end{aligned}$$

Let $\mathbb{F} = (\mathcal{F}_t)_{0 \leq t \leq T}$ denote the right-continuous augmentation of the filtration generated by ξ, B^1, B^2 and B^0 , and let $\mathbb{F}^0 = (\mathcal{F}_t^0)_{0 \leq t \leq T}$ denote the right continuous augmentation of the filtration generated by B^0 ². For $t_0 \in [0, T]$ and $m_0 \in \mathcal{P}(\mathbb{T}^{2d})$ fixed, we let $\mathcal{A}_{t_0, m_0}$ be the set of A -valued processes $\alpha = (\alpha)_{t_0 \leq t \leq T}$, progressively measurable with respect to $\mathbb{F}$. Then,

$$\begin{aligned}U^\varepsilon(t_0, m_0) &= \inf_{\alpha \in \mathcal{A}_{t_0, m_0}} \mathbb{E}^{\mathbb{P}^{m_0}} \left[\int_{t_0}^T L \left(X_s^{t_0, \varepsilon, m_0}, Y_s^{t_0, \varepsilon, m_0}, \alpha_s \right) ds \right. \\ &\quad \left. + \int_{t_0}^T \mathcal{F}(\mathcal{L}^{0, m_0}(X_s^{t_0, \varepsilon, m_0})) ds + \mathcal{G}(\mathcal{L}^{0, m_0}(X_T^{t_0, \varepsilon, m_0})) \right],\end{aligned}\tag{3.1.6}$$

where for each $\alpha \in \mathcal{A}_{t_0, m_0}$, the processes $X_t^{t_0, \varepsilon, m_0}, Y_t^{t_0, \varepsilon, m_0}$ solve the system of SDEs ($t \geq t_0$)

$$\begin{cases} dX_t^\varepsilon &= c(X_t^\varepsilon, Y_t^\varepsilon, \alpha_t) dt + \sqrt{2}\sigma_1 dB_t^1 + \sqrt{2}\sigma_{0,1} dB_t^0, \\ dY_t^\varepsilon &= \frac{1}{\varepsilon} f(X_t^\varepsilon, Y_t^\varepsilon, \alpha_t) dt + \frac{\sqrt{2}}{\sqrt{\varepsilon}} \sigma_2 dB_t^2 + \frac{\sqrt{2}}{\sqrt{\varepsilon}} \sigma_{0,2} dB_t^0, \end{cases}\tag{3.1.7}$$

with initial condition $\mathcal{L}((X_{t_0}^\varepsilon, Y_{t_0}^\varepsilon)) = m_0$ and where for a random variable X , $\mathcal{L}^{0, m_0}(X)$ is the law of the $\mathcal{F}_s$ -measurable random variable X conditional on $\mathcal{F}_s^0$ with respect to $\mathbb{P}^{m_0}$.

The above system (3.1.7) is a system of slow-fast SDEs (X is called the slow variable and

2. B^1, B^2 play the role of the idiosyncratic noise, while B^0 plays the role of the common noise.

Y is called the fast variable). The study of the behavior of such systems as ε approaches 0 is a topic of intense research, especially in the absence of controls (c, f do not depend on α). We cite the initial works [PV01b] and [RX21b], which were later generalized in [RSX19a, LWX22b, QW22b, HLS25, BS23a, HLX24], which treat the case of $\sigma_{0,1} = \sigma_{0,2} = \mathbb{O}_{d \times d}$. The case of $\sigma_{0,1}, \sigma_{0,2} \neq \mathbb{O}_{d \times d}$ was studied in [Zit23]. Similar problems were also studied in [BK23, BB05, Kum12], in different settings and for initial conditions in (3.1.7) in $\mathbb{R}^{2d}$, that is $(X_{t_0}^\varepsilon, Y_{t_0}^\varepsilon) = (x, y) \in \mathbb{R}^{2d}$. Listing all the related literature is beyond the scope of this paper.

With the above terminology, our result gives a characterization of the behavior of the value function of a *slow-fast stochastic optimal control problem*, as $\varepsilon \rightarrow 0$. If we restrict to the case where $L = 0$, $\mathcal{F} = 0$ and the functions f, c do not depend on α , we may recover the homogenization result for slow-fast SDEs from [Zit23, Remark 2.3(vii), Theorem 2.1] in the case where the intensity of the common noise B^0 is not state dependent. We state it as a Corollary.

Corollary 3.1.3. *Assume (A) from section 2 in the particular case where $L = 0$, $\mathcal{F} = 0$ and c, f are independent of α . We also consider $T \geq t_0 \geq 0$ and $(X_t^{t_0, \varepsilon, m_0}, Y_t^{t_0, \varepsilon, m_0})$ to be a solution of (3.1.7) with initial condition on the law $\mathcal{L}(X_{t_0}^{t_0, \varepsilon, m_0}, Y_{t_0}^{t_0, \varepsilon, m_0}) = m_0$. Then, $\mathcal{G}(\mathcal{L}(X_T^{t_0, \varepsilon, m_0})) \xrightarrow{\varepsilon \rightarrow 0} \mathcal{G}(\mathcal{L}(X_T^{t_0, \pi_*^1 m_0}))$, where $X_t^{t_0, \pi_*^1 m_0}$ is the process satisfying the SDE*

$$dX_t = \bar{c}(X_t)dt + \sqrt{2}\sigma_1 dB'_t + \sqrt{2}\sigma_{0,1} dB_t^0, \quad \mathcal{L}(X_{t_0}) = \pi_*^1 m_0, \quad (3.1.8)$$

B' is a Brownian motion on a possibly different probability space and $\bar{c} : \mathbb{T}^d \rightarrow \mathbb{R}^d$ is a function that depends on c, f, σ_2 and $\sigma_{0,2}$.

3.1.3 Extensions and Outlook

Our proof of Theorem 3.1.1 without difficulty to the case where $\sigma_1 := \sigma_1(x), \sigma_2 := \sigma_2(x, y)$ are also state dependent, assuming sufficient regularity. If we further suppose that c, f, L, σ_1 and σ_2 also depend on the law of the slow variable (X_t^ε) , then the Hamiltonian is no longer separated³, that is it is a general function $H(x, y, p, q, \mu) : \mathbb{T}^{2d} \times \mathbb{R}^{2d} \times \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$. In that case, equation (3.1.1) now takes the form

$$\begin{cases} -\partial_t U^\varepsilon + \int_{\mathbb{T}^{2d}} H^\varepsilon(x, y, D_m U^\varepsilon(t, m, x, y), \pi_*^1 m) m(dx, dy) \\ \quad - \int_{\mathbb{T}^{2d}} \text{tr} \left((\sigma^\varepsilon \sigma^{\varepsilon\top} + \sigma_0^\varepsilon \sigma_0^{\varepsilon\top}) D_{(x,y)} D_m U^\varepsilon(t, m, x, y) \right) m(dx, dy) \\ \quad - \int_{\mathbb{T}^{2d}} \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_0^\varepsilon \sigma_0^{\varepsilon\top} D_{mm}^2 U^\varepsilon(t, m, x, y, x', y') \right) m(dx, dy) m(dx', dy') = 0, \\ U^\varepsilon(T, m) = \mathcal{G}(\pi_*^1 m). \end{cases}$$

Assuming Lipschitz continuity on H, c, f, σ_2 and L in the measure variable, our proof carries through with the suitable modifications in Proposition 3.4.1 to get the additional Lipschitz regularity of $\bar{H}(x, p, \mu)$ in the measure variable. In this case, Corollary 3.1.3 yields a homogenization result for conditional McKean-Vlasov slow-fast SDEs. We omit these more general cases to simplify the presentation.

Our proof does not work when $\sigma_{0,1}, \sigma_{0,2}$ are state dependent. The main difficulty is the comparison principle for (3.1.4), which is necessary in order to establish that the upper and lower semi-limits are equal (see proof of Theorem 3.1.1). For a general second order Hamilton-Jacobi equation in the Wasserstein space a comparison principle for viscosity sub/super-solutions has been proved in [BCE⁺25], provided that the Hamiltonian is convex. The perturbed test function argument would still work if we adopt the definition from [BCE⁺25],

3. This means that it takes the form $H(x, y, p, q) - \mathcal{F}(\mu)$.

however $\overline{H}$ has to satisfy the condition:

$$\overline{H}(x, p) = \sup_{\alpha \in A} \{-b(x, \alpha) \cdot p - L(x, \alpha)\}, \quad (3.1.9)$$

for some $b, L : \mathbb{T}^d \times A \rightarrow \mathbb{R}$ jointly continuous and Lipschitz continuous. If (3.1.9) does not hold, it is unclear whether there is comparison for (3.1.4).

The following natural directions, where there is comparison principle for viscosity subsolutions and supersolutions, also remain unanswered:

- Fully non-linear and convex in p, q, X, Y Hamiltonians $H(x, y, p, q, X, Y)$ (see [CGK⁺24, CTQ25]). In this case a different analysis is needed at the level of Proposition 3.4.1, as well as a representation for the effective Hamiltonian analogous to (3.1.9).
- The function $\mathcal{G}$ in (3.1.1) depends on the full measure m instead of just the first marginal $\pi_*^1 m$. The finite dimensional analogue of this would be for g in (3.1.2) to depend on both x, y and was studied in [AB03] (and references therein).

We leave these questions for future research.

We finish the introduction with the organization of the paper

3.1.4 Organization of the chapter.

Section 2 introduces the notation and the assumptions used throughout the paper. In section 3 we provide the definition of viscosity solutions and state basic properties already existing in the literature. In section 4 we give the formal idea of the proof, which is the perturbed function method adapted to the infinite dimensional setting, and the proof of the main results. We leave the definitions and the properties of the Wasserstein space and the Wasserstein derivatives as well as the proofs of some technical lemmas for the last two sections.

3.2 Notation and Assumptions

We now introduce the notations and assumptions used throughout the rest of the paper.

Notation. Through the note $T > 0$. We will be using the symbol $\mathbf{x} := (x, y) \in \mathbb{T}^d \times \mathbb{T}^d$. For a Polish space X and $p \geq 1$, we will denote the space of probability measures over X with finite p -moment by $\mathcal{P}_p(X)$ and by $\mathbf{d}_p$ its distance; we refer to section 3.5 for the precise definitions. We use the notation $\mathcal{P}(X) := \mathcal{P}_1(X)$. Differentiation of functions defined in this space is explained in section 3.5.

Let $m \in \mathcal{P}_p(\mathbb{T}^{2d})$ and $\pi^1 : \mathbb{T}^d \times \mathbb{T}^d \rightarrow \mathbb{T}^d$ with $\pi^1(x, y) = x$. We denote by $\pi_*^1 m \in \mathcal{P}_p(\mathbb{T}^d)$ the x -marginal of the measure m . If $f : \mathbb{T}^d \rightarrow \mathbb{T}^d$ and $\mu \in \mathcal{P}(\mathbb{T}^d)$, then $f_*\mu \in \mathcal{P}(\mathbb{T}^d)$ is the pushforward measure such that $f_*\mu(B) = \mu(f^{-1}(B))$ for any measurable $B \subset \mathbb{T}^d$.

We use the symbol $D_{(x,y)}u(x, y) = (D_x u, D_y)$ for the gradient as a vector of the gradients. $u \in C^k$ means that the function is k times continuously differentiable. For a function $u(t, x)$ we write $u \in C^{1,2}$ if the derivatives $D_t u$, $D_x u$, $D_{xx}^2 u$ exist and are continuous. We also use the notation $u \in C_x^2$ (resp. C_x^1) if the derivatives $D_x u$ and $D_{xx}^2 u$ (resp. $D_x u$) exist and are continuous with respect to the x variable. We write $u \in C^{2,\beta}$ if u is twice continuously differentiable and the second derivative is β -Hölder continuous. We use the usual Hölder norm. We denote by $f * g(x) := \int f(y)g(x - y)dy$ the convolution between two functions. $\mathbb{I}_{d \times d}$ is the d -dimensional unit matrix and $\mathbb{O}_{d \times d}$ the d -dimensional zero matrix.

Assumptions. We work under the following assumptions on the functions H , $\mathcal{F}$ and $\mathcal{G}$. Below, A is compact subset of $\mathbb{R}^{2d}$.

(A1) $c, f : \mathbb{T}^d \times \mathbb{T}^d \times A \rightarrow \mathbb{R}^d$ are continuous and there exists a constant $K \geq 0$ such that

$$|c(x, y, \alpha) - c(x', y', \alpha)| + |f(x, y, \alpha) - f(x', y', \alpha)| \leq K (|x - x'| + |y - y'|),$$

for all $\alpha \in A$, $x, x', y, y' \in \mathbb{T}^d$, where $|x - x'|$ denotes the Euclidean norm of $x - x'$.

(A2) The Lagrangian $L : \mathbb{T}^d \times \mathbb{T}^d \times A \rightarrow \mathbb{R}$ is continuous and satisfies

$$|L(x, y, a) - L(x', y', a)| \leq C_L (|x - x'| + |y - y'|),$$

for some constant C_L , for all $\alpha \in A$ and $x, x', y, y' \in \mathbb{T}^d$.

(A3) The Hamiltonian $H : \mathbb{T}^{2d} \times \mathbb{R}^{2d} \rightarrow \mathbb{R}$ satisfies for some constants $C_H > 0$ and $\beta \in (0, 1)$:

(1) H satisfies (3.1.5) with c, f, L as in assumptions (A1) and (A2).

(2) For any $\mathbf{x}, \mathbf{x}' \in \mathbb{T}^{2d}$ and any $p, q, p', q' \in \mathbb{R}^d$ we have

$$\begin{aligned} & |H(\mathbf{x}, p, q) - H(\mathbf{x}', p', q')| \\ & \leq C_H(1 + |p| + |q| + |p'| + |q'|)(|\mathbf{x} - \mathbf{x}'| + |p - p'| + |q - q'|), \end{aligned} \tag{3.2.1}$$

(3) H is a C^2 function with the second derivative being locally β -Hölder continuous.

Furthermore, for any $x, y \in \mathbb{T}^d$ and $p, q \in \mathbb{R}^d$

$$|D_x H(x, y, p, q)| \leq C_H(1 + |p|), \tag{3.2.2}$$

$$|D_p H(x, y, p, q)| \leq C_H(1 + |p|). \tag{3.2.3}$$

(A4) $\mathcal{F}, \mathcal{G} : \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$ are $\mathbf{d}_1$ -Lipschitz continuous, i.e there exists a constant C_{FG} such

that

$$|\mathcal{F}(\mu) - \mathcal{F}(\mu')| + |\mathcal{G}(\mu) - \mathcal{G}(\mu')| \leq C_{FG} \mathbf{d}_1(\mu, \mu'), \text{ for every } \mu, \mu' \in \mathcal{P}(\mathbb{T}^d).$$

(A5) $\sigma_1, \sigma_2, \sigma_{0,1}, \sigma_{0,2}$ are $d \times d$ matrices and σ_1, σ_2 are invertible.

We will denote the set of assumption (A1)-(A5) as assumption **(A)**.

Remark 3.2.1. (i) Assumptions (A4), (A5) and (A3)(2) guarantee the existence of a viscosity solution to (3.1.1) (see section 3.3).

(ii) There are many examples of Hamiltonians H that satisfy (A3). Three of them are

- $H \equiv 0$.
- $H(x, y, p, q) = p \cdot c(x, y) + q \cdot f(x, y)$ with $L \equiv 0$ and f, d independent of α .
- $L(x, y, \alpha) = \frac{|\alpha|^2}{2}$, $c(x, y, \alpha) = \alpha$, $f(x, y, \alpha) = \alpha$, H satisfies (3.1.5) and A is closed ball centered at the origin.

(iii) Due to the compactness of $\mathcal{P}(\mathbb{T}^d)$, assumption (A4) also implies that $\mathcal{F}, \mathcal{G}$ are bounded.

3.3 On viscosity solutions

To provide more generality to our statements, for this section only we change the notation, but we keep the same symbols for the Hamiltonian, the running cost $\mathcal{F}$ and the final cost $\mathcal{G}$. Let $n \in \mathbb{N}^*$, $H : \mathbb{T}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$, $\mathcal{F}, \mathcal{G} : \mathcal{P}(\mathbb{T}^n) \rightarrow \mathbb{R}$, an $n \times n$ invertible matrix σ and $n \times n$

matrix σ_0 . We consider the Hamilton-Jacobi equation posed in $[0, T] \times \mathcal{P}(\mathbb{T}^n)$

$$\left\{ \begin{array}{l} -\partial_t U(t, m) + \int_{\mathbb{T}^n} H(x, D_m U(t, m, x)) m(dx) \\ \quad - \int_{\mathbb{T}^n} \text{tr} \left((\sigma \sigma^\top + \sigma_0 \sigma_0^\top) D_x D_m U(t, m, x) \right) m(dx) \\ \quad - \int_{\mathbb{T}^n} \int_{\mathbb{T}^n} \text{tr} \left(\sigma_0 \sigma_0^\top D_{mm}^2 U(t, m, x, y) \right) m(dx) m(dy) = \mathcal{F}(m), \\ U(T, m) = \mathcal{G}(m). \end{array} \right. \quad (3.3.1)$$

In this section we introduce the notion of viscosity solutions of HJ equations that we will be using as well as their properties. Before writing the definition we note the observation from [BEZ23] that if U is smooth and we define $\tilde{U} : [0, T] \times \mathbb{T}^n \times \mathcal{P}(\mathbb{T}^n) \rightarrow \mathbb{R}$ via the formula

$$\tilde{U}(t, z, m) = U(t, (\text{Id} + z)_* m), \quad (3.3.2)$$

then we have

$$\begin{aligned} \text{tr} \left(\sigma_0 \sigma_0^\top D_{zz}^2 \tilde{U}(t, z, m) \right) &= \int_{\mathbb{T}^n} \text{tr} \left(\sigma_0 \sigma_0^\top D_x D_m U(t, m^z, x) \right) m^z(dx) \\ &\quad + \int_{\mathbb{T}^n} \int_{\mathbb{T}^n} \text{tr} \left(\sigma_0 \sigma_0^\top D_{mm}^2 U(t, m^z, x, y) \right) m^x(dx) m^z(dy), \end{aligned} \quad (3.3.3)$$

where

$$m^z = (\text{Id} + z)_* m. \quad (3.3.4)$$

Using this observation and (3.3.1), one can check that $\tilde{U}$ satisfies

$$\left\{ \begin{array}{l} -\partial_t \tilde{U} + \int_{\mathbb{T}^n} \tilde{H}(x, z, D_m \tilde{U}(t, z, m, x)) m(dx) \\ \quad - \int_{\mathbb{T}^n} \text{tr} \left(\sigma \sigma^\top D_x D_m \tilde{U}(t, z, m, x) \right) m(dx) \\ \quad - \text{tr} \left(\sigma_0 \sigma_0^\top D_{zz}^2 \tilde{U}(t, z, m, x, y) \right) m(dx) m(dy) = \tilde{\mathcal{F}}(z, m), \\ \tilde{U}(T, z, m) = \tilde{\mathcal{G}}(z, m), \end{array} \right. \quad (3.3.5)$$

for $(t, z, m) \in [0, T] \times \mathbb{T}^n \times \mathcal{P}(\mathbb{T}^n)$, where

$$\tilde{H}(x, z, p) := H(x + z, p), \quad \tilde{\mathcal{F}}(z, m) := \mathcal{F}(m^z), \quad \tilde{\mathcal{G}}(z, m) := \mathcal{G}(m^z). \quad (3.3.6)$$

We define viscosity solutions for the original equation (3.3.1) in terms of viscosity solutions of the transformed equation (3.3.5). The latter is much easier to deal with, because the terms involving second-order derivatives in measure have been exchanged for a finite dimensional operator (3.3.3). Viscosity solutions for (3.3.5) are defined in the sense of test functions, in the spirit of Crandall-Lions.

Definition 3.3.1. We say that $\varphi : [0, T] \times \mathbb{T}^n \times \mathcal{P}(\mathbb{T}^n) \rightarrow \mathbb{R}$ is a test function and we write $\varphi \in C^{1,2,2}$, if $\partial_t \varphi(t, z, m)$, $D_m \varphi(t, z, m, x)$, $D_x D_m \varphi(t, z, m, x)$, $D_{mm}^2 \varphi(t, z, m, x, y)$ exist and are continuous as well as $D_z \varphi(t, z, m)$, $D_{zz}^2 \varphi(t, z, m)$ exist and are continuous.

Remark 3.3.2. We keep the above definition for test functions $\varphi : [0, T] \times \mathbb{T}^n \times \mathbb{T}^n \times \mathcal{P}(\mathbb{T}^n \times \mathbb{T}^n) \rightarrow \mathbb{R}$, where $\mathbb{T}^n$ is substituted with $\mathbb{T}^n \times \mathbb{T}^n$ and $\mathcal{P}(\mathbb{T}^n)$ with $\mathcal{P}(\mathbb{T}^n \times \mathbb{T}^n)$. We denote the set of such test functions by $C^{1,2,2}$ again.

Having this in mind, we introduce the notion of viscosity solutions. We will be using the same definition of viscosity solutions (with the obvious modifications) for (3.1.1) with $2d$ in place of n .

Definition 3.3.3. An upper (resp. lower) semi-continuous function $\tilde{U} : [0, T] \times \mathbb{T}^n \times \mathcal{P}(\mathbb{T}^n) \rightarrow \mathbb{R}$ is a viscosity subsolution (resp. supersolution) to equation (3.3.5) if:

- $\tilde{U}(T, z, m) \leq$ (resp. $\geq$) $\tilde{\mathcal{G}}(z, m)$, for every $(z, m) \in \mathbb{T}^n \times \mathcal{P}(\mathbb{T}^n)$;
- for every $(t, z, m) \in [0, T)$ and any test function $\varphi \in C^{1,2,2}$ such that $\tilde{U} - \varphi$ has a maximum (resp. minimum) at (t, z, m) , the first equation in (3.3.5) is satisfied with the inequality $\leq$ (resp. $\geq$) instead of the equality and φ in place of $\tilde{U}$.

Finally, $\tilde{U}$ is a viscosity solution of (3.3.5) if it is both a viscosity subsolution and a viscosity supersolution.

Remark 3.3.4. Using a negative Sobolev space argument as in [DJS23, Lemma 2.3], we can infer that just like in finite dimensions, in order to check that a function U is a viscosity sub/super-solution to (3.3.5), it suffices to restrict our attention to strict local maxima/minima. We may also assume that the strict local maximum/minimum is 0 by shifting the test function if necessary.

Definition 3.3.5. An upper (resp. lower) semicontinuous function $U : [0, T] \times \mathcal{P}(\mathbb{T}^n) \rightarrow \mathbb{R}$ is called a viscosity subsolution (resp. supersolution) of (3.3.1) if $\tilde{U} : [0, T] \times \mathbb{T}^n \times \mathcal{P}(\mathbb{T}^n) \rightarrow \mathbb{R}$ defined as in (3.3.2) is a viscosity subsolution (resp. supersolution) of (3.3.5). U is a viscosity solution of (3.3.1) if it is a viscosity subsolution and a viscosity supersolution.

We note that from [DJS23, Theorem 2.6, Theorem 2.7]⁴, we have the following result for viscosity solutions of (3.3.1).

Theorem 3.3.6. *Suppose that Assumption (A3)(2) for $n = 2d$ holds and that $\mathcal{F}, \mathcal{G} : \mathcal{P}(\mathbb{T}^n) \rightarrow \mathbb{R}$ are Lipschitz continuous. Then, there exists a viscosity solution U of (3.3.1). Furthermore, if U_1 is a viscosity subsolution of (3.3.1) and U_2 a viscosity supersolution of (3.3.1), then $U_1 \leq U_2$ in $[0, T] \times \mathcal{P}(\mathbb{T}^n)$. In particular, there is at most one viscosity solution of (3.3.1).*

Due to the analysis done at the beginning of the section, if $\sigma_{0,1} = \sigma_{0,2} = \mathbb{O}_{d \times d}$, then there is no need to introduce (3.3.5) into the definition of viscosity solutions. Thus, Definition 3.3.5 will be in the sense of test functions as in the classical definition of Crandall-Lions in finite dimensions. For completeness, we include the proof of Theorem 3.3.6 in the simpler case

4. This theorem is true for much more general assumptions than **(A)**, in particular we only need (A3)(2). It is also true even when H depends on the measure variable as long as H is locally Lipschitz in the measure variable.

where Assumption (A3) is in place (for $n = 2d$) and the following assumptions **(A')** hold in section 3.6 (Proposition 3.6.1 for the existence and Proposition 3.6.3 for the uniqueness).

$$\left\{ \begin{array}{l} \text{(i) } \sigma_0 = 0, \text{ that is, there is no common noise. } \sigma \text{ is an invertible } n \times n \text{ matrix.} \\ \text{(ii) } H \text{ is strictly convex and } \mathcal{F}, \mathcal{G} \text{ are } C^3 \text{ in } \mathcal{P}(\mathbb{T}^n). \\ \text{(iii) There is a bounded continuous function } b : \mathbb{T}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n, \text{ which is Lipschitz in the} \\ x \text{ argument, such that the function } L : \mathbb{T}^n \times A \rightarrow \mathbb{R} \\ \text{with } L(x, \alpha) = \sup_{p \in \mathbb{R}^n} \{-b(x, \alpha) \cdot p - H(x, p)\} \text{ is bounded, Lipschitz and } C^2. \end{array} \right. \quad (\mathbf{A}')$$

A Corollary of Theorem 3.3.6 is the following.

Corollary 3.3.7. *Assume (A3)(2) and (A4) for $2d = n$ and let U be the unique viscosity solution of (3.3.1). Then, there exists a constant $C > 0$ depending only on $\max_{\mu} |\mathcal{F}(\mu)|$, $\max_{\mu} |\mathcal{G}(\mu)|$ and $\|H(\cdot, 0)\|_{L^\infty}$ such that $|U(t, m)| \leq C$ for every $t \in [0, T]$ and $m \in \mathcal{P}(\mathbb{T}^n)$.*

Proof. We will prove that there exists a constant $C > 0$ depending only on $\mathcal{F}, \mathcal{G}$ and $\|H(\cdot, 0)\|_{L^\infty}$ such that $-C \leq \tilde{U}(t, z, m) \leq C$ for all $(t, z, m) \in [0, T] \times \mathbb{T}^n \times \mathcal{P}(\mathbb{T}^n)$, where $\tilde{U}$ is as in (3.3.2). We start with the upper bound.

Let C_1, C_2 be two numbers that will be chosen later and assume that $\tilde{U}(t, z, m) + C_1 t + C_2$ attains its maximum at $(t_0, z_0, m_0) \in [0, T] \times \mathbb{T}^n \times \mathcal{P}(\mathbb{T}^n)$. The choice of C_2 is made such that the maximum of $\tilde{U}(t, z, m) + C_1 t + C_2$ is 0. There are two cases: either $t_0 \in [0, T]$ or $t_0 = T$.

In the first case, by the definition of viscosity solutions for (3.3.5) with test function $-C_1 t - C_2$ we get

$$C_1 + \int_{\mathbb{T}^n} \tilde{H}(x, z_0, 0) m(dx) \leq \tilde{\mathcal{F}}(z_0, m_0),$$

thus $C_1 \leq \|H(\cdot, 0)\|_{L^\infty} + \max_{m \in \mathcal{P}(\mathbb{T}^n)} \mathcal{F}(m)$. This is a contradiction if C_1 is chosen to be

large enough, therefore we may assume that $t_0 = T$. We have

$$\tilde{U}(t, z, m) + C_1 t + C_2 \leq \tilde{\mathcal{G}}(z_0, m_0) + C_1 T + C_2, \text{ for all } (t, z, m) \in [0, T] \times \mathbb{T}^n \times \mathcal{P}(\mathbb{T}^n).$$

So, $\tilde{U}(t, z, m) \leq \max_{m \in \mathcal{P}(\mathbb{T}^n)} \mathcal{G}(m) + |C_1|T$. The upper bound follows. To prove the lower bound we use a similar argument for the minimum of $U(t, m) - C_1 t - C_2$. $\square$

3.4 Proof of the main result

We start by presenting a formal calculation that motivates our argument.

3.4.1 Formal calculation

Let's assume that $U^\varepsilon(t, m) = U(t, \mu) + \varepsilon V(t, m)$, where μ is the x -marginal of m ($\mu = \pi_*^1 m$) and U, V are two regular functions (to be chosen later). The idea is that U^ε will be an ε -perturbation of its limit U . Substituting into (3.1.1) and using Proposition 3.5.5 yields

$$\begin{aligned} & -\partial_t U(t, \mu) - \varepsilon \partial_t V(t, m) \\ & + \int_{\mathbb{T}^{2d}} H \left(x, y, D_\mu U(t, \mu, x) + \varepsilon D_x \frac{\delta V}{\delta m}(t, m, x, y), D_y \frac{\delta V}{\delta m}(t, m, x, y) \right) m(dx, dy) \\ & - \int_{\mathbb{T}^d} \text{tr} \left(\sigma_1 \sigma_1^\top D_x D_\mu U(t, \mu, x) \right) \mu(dx) \\ & - \varepsilon \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_2 \sigma_2^\top D_{xx}^2 \frac{\delta V}{\delta m}(t, m, x, y) \right) m(dx, dx) \\ & - \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_2 \sigma_2^\top D_{yy}^2 \frac{\delta V}{\delta m}(t, m, x, y) \right) m(dx, dx) \\ & - \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_{0,1} \sigma_{0,1}^\top D_x D_\mu U(t, \mu, x) \right) \mu(dx) \\ & - \varepsilon \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_0^\varepsilon \sigma_0^{\varepsilon\top} D_{(x,y)} D_m V(t, m, x, y) \right) m(dx, dy) \\ & - \int_{\mathbb{T}^{2d}} \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_{0,1} \sigma_{0,1}^\top D_{\mu\mu}^2 U(t, \mu, x, x') \right) \mu(dx) \mu(dx') \end{aligned}$$

$$- \varepsilon \int_{\mathbb{T}^{2d}} \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_0^\varepsilon \sigma_0^{\varepsilon^\top} D_{mm}^2 V(t, m, x, y, x', y') \right) m(dx, dy) m(dx', dy') = \mathcal{F}(\mu).$$

By the regularity on H , the third term is equal to

$$\int_{\mathbb{T}^{2d}} H \left(x, y, D_\mu U(t, \mu, x), D_y \frac{\delta V}{\delta m}(t, m, x, y) \right) m(dx, dy) + \mathcal{O}(\varepsilon),$$

therefore collecting the $\mathcal{O}(\varepsilon)$ terms, the equation becomes

$$\begin{aligned} \mathcal{O}(\varepsilon) + \mathcal{F}(\mu) = & -\partial_t U(t, \mu) + \int_{\mathbb{T}^{2d}} H \left(x, y, D_\mu U(t, \mu, x), D_y \frac{\delta V}{\delta m}(t, m, x, y) \right) m(dx, dy) \\ & - \int_{\mathbb{T}^d} \text{tr} \left(\sigma_1 \sigma_1^\top D_x D_\mu U(t, \mu, x) \right) \mu(dx) \\ & - \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_2 \sigma_2^\top D_{yy}^2 \frac{\delta V}{\delta m}(t, m, x, y) \right) m(dx, dy) \\ & - \int_{\mathbb{T}^{2d}} \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_{0,1} \sigma_{0,1}^\top D_{\mu\mu}^2 U(t, \mu, x, x') \right) \mu(dx) \mu(dx') \\ & - \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_{0,1} \sigma_{0,1}^\top D_x D_\mu U(t, \mu, x) \right) \mu(dx) \\ & - \int_{\mathbb{T}^{2d}} \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_{0,2} \sigma_{0,2}^\top D_{yy'}^2 \frac{\delta^2 V}{\delta m^2}(t, m, x, y, x', y') \right) m(dx, dy) m(dx', dy') \\ & - \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_{0,2} \sigma_{0,2}^\top D_{yy}^2 \frac{\delta V}{\delta m}(t, m, x, y) \right) m(dx, dy). \end{aligned} \quad (3.4.1)$$

As U is expected to be the limit of U^ε as $\varepsilon \rightarrow 0$, to find the equation satisfied by it we need to eliminate V from (3.4.1). To do that, we would like to have

$$\begin{aligned} & \int_{\mathbb{T}^{2d}} H \left(x, y, D_\mu U(t, \mu, x), D_y \frac{\delta V}{\delta m}(t, m, x, y) \right) m(dx, dy) \\ & - \int_{\mathbb{T}^{2d}} \text{tr} \left((\sigma_2 \sigma_2^\top + \sigma_{0,2} \sigma_{0,2}^\top) D_{yy}^2 \frac{\delta V}{\delta m}(t, m, x, y) \right) m(dx, dy) \\ & - \int_{\mathbb{T}^{2d}} \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_{0,2} \sigma_{0,2}^\top D_{yy'}^2 \frac{\delta^2 V}{\delta m^2}(t, m, x, y, x', y') \right) m(dx, dy) m(dx', dy') = B(\mu), \end{aligned}$$

for some function $B : \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$ depending on U . This leads us to considering the problem

$$\begin{aligned} \overline{H}(x, D_\mu U(t, \mu, x)) &= H(x, y, D_\mu U(t, \mu, x), D_y \phi(t, \mu, x, y)) \\ &\quad - \operatorname{tr} \left((\sigma_2 \sigma_2^\top + \sigma_{0,2} \sigma_{0,2}^\top) D_{yy}^2 \phi(t, \mu, x, y) \right), \end{aligned} \quad (3.4.2)$$

where $\overline{H}$ is the unique function such that this problem has a unique solution ϕ . Since we want $D_y \phi = D_y \frac{\delta V}{\delta m}$, we let

$$V(t, m) = \int_{\mathbb{T}^{2d}} \phi(t, \mu, x, y) m(dx, dy).$$

Then, again by Proposition 3.5.5,

$$\frac{\delta V}{\delta m}(t, m, x, y) = \phi(t, \mu, x, y) + \int_{\mathbb{T}^{2d}} \frac{\delta \phi}{\delta \mu}(t, \mu, x', y', x) m(dx', dy'),$$

so $D_y \frac{\delta V}{\delta m}(t, m, x, y) = D_y \phi(t, \mu, x, y)$, which implies,

$$\begin{aligned} &\int_{\mathbb{T}^{2d}} H \left(x, y, D_\mu U(t, \mu, x), D_y \frac{\delta V}{\delta m}(t, m, x, y) \right) m(dx, dy) \\ &\quad - \int_{\mathbb{T}^{2d}} \operatorname{tr} \left((\sigma_2 \sigma_2^\top + \sigma_{0,2} \sigma_{0,2}^\top) D_{yy}^2 \frac{\delta V}{\delta m}(t, m, x, y) \right) m(dx, dy) \\ &\quad - \int_{\mathbb{T}^{2d}} \int_{\mathbb{T}^{2d}} \operatorname{tr} \left(\sigma_{0,2} \sigma_{0,2}^\top D_{yy'}^2 \frac{\delta^2 V}{\delta m^2}(t, m, x, y, x', y') \right) m(dx, dy) m(dx', dy') \\ &\quad = \int_{\mathbb{T}^{2d}} \overline{H}(x, D_\mu U(t, \mu, x)) \mu(dx), \end{aligned}$$

due to (3.4.2) and the fact that $\frac{\delta^2 V}{\delta m^2}$ does not have a term depending on both y and y' .

Therefore, for that choice of V (3.4.1), gives

$$\begin{aligned} &-\partial_t U(t, \mu) + \int_{\mathbb{T}^{2d}} \overline{H}(x, D_\mu U(t, \mu, x)) \mu(dx) \\ &\quad - \int_{\mathbb{T}^d} \operatorname{tr} \left((\sigma_1 \sigma_1^\top + \sigma_{0,1} \sigma_{0,1}^\top) D_x D_\mu U(t, \mu, x) \right) \mu(dx) \end{aligned}$$

$$- \int_{\mathbb{T}^d} \int_{\mathbb{T}^d} \text{tr} \left(\sigma_{0,1} \sigma_{0,1}^\top D_{\mu\mu}^2 U(t, \mu, x, x') \right) \mu(dx) \mu(dx') = \mathcal{O}(\varepsilon) + \mathcal{F}(\mu).$$

We send $\varepsilon \rightarrow 0$ to derive that U satisfies equation (3.1.4).

3.4.2 Proof of Theorem 3.1.1

As the formal calculation suggests, for $(x, p) \in \mathbb{T}^d \times \mathbb{R}^d$, to find $\bar{H}$ we need to study the following problem, which is posed in $\mathbb{T}^d$:

$$H(x, y, p, D_y v) - \text{tr} \left((\sigma_2 \sigma_2^\top + \sigma_{0,2} \sigma_{0,2}^\top) D_{yy}^2 v \right) = l.$$

Such problems have been studied in the context of homogenization theory, for example in [Eva92], and it is known that there exists a unique constant l such that (3.4.2) is well posed (up to an additive constant). We provide an argument on the existence of such a constant and we show some properties that will be useful in the proof of the main Theorem 3.1.1. In particular, the following Proposition holds.

Proposition 3.4.1. *Assume that H satisfies (A3). Then, for any $\delta > 0$ and $x \in \mathbb{T}^d$, $p \in \mathbb{R}^d$, there exists a unique classical solution $v^\delta := v^\delta(x, p, y)$ of*

$$\delta v^\delta - \text{tr} \left((\sigma_2 \sigma_2^\top + \sigma_{0,2} \sigma_{0,2}^\top) D_{yy}^2 v^\delta \right) + H(x, y, p, D_y v^\delta) = 0, \quad y \in \mathbb{T}^d. \quad (3.4.3)$$

The function $v^\delta(x, p, y) \in C_{x,p,y}^2$, that is the derivatives $D_x v^\delta, D_{xx}^2 v^\delta, D_p v^\delta, D_{pp}^2 v^\delta, D_{xp}^2 v^\delta$ exist and are continuous, and for any $(x, p) \in \mathbb{T}^d \times \mathbb{R}^d$ there exists a number $\bar{H}(x, p)$ such that

$$\lim_{\delta \rightarrow 0} \left\| \delta v^\delta(x, p, \cdot) + \bar{H}(x, p) \right\|_{L^\infty} = 0. \quad (3.4.4)$$

Furthermore, the function $\overline{H} : \mathbb{T}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ satisfies for some constant $C > 0$

$$\begin{aligned} & |\overline{H}(x, p) - \overline{H}(x', p')| \\ & \leq C(1 + |p| + |p'|)(|x - x'| + |p - p'|), \text{ for all } x, x' \in \mathbb{T}^d, p, p' \in \mathbb{R}^d. \end{aligned} \quad (3.4.5)$$

Proof. Throught the proof, we use the notation $a = \sigma_2 \sigma_2^\top + \sigma_{0,2} \sigma_{0,2}^\top$, which is a strictly positive matrix due to assumption (A5). The existence and the uniqueness of a classical solution $v^\delta(x, p, \cdot)$, where $x \in \mathbb{T}^d$, $p \in \mathbb{R}^d$ are fixed, is a standard result and can be found in [GT77]. It suffices to show the other statements. We will divide the proof in 4 steps.

Step 1. (estimates on v^δ and existence of $\overline{H}(x, p)$)

Let $\lambda(p) = \|H(\cdot, \cdot, p, 0)\|_{L^\infty}$. We will show that there exists $\beta \in (0, 1)$ and a constant $C > 0$ independent of x, p such that

$$\left\| v^\delta(x, p, \cdot) - v^\delta(x, p, 0) \right\|_{C^{2,\beta}} \leq C(1 + |p| + \lambda(p)). \quad (3.4.6)$$

Our argument follows that in [AL98], so we provide the outline. Considering the optimality conditions for a maximum or a minimum of v^δ in (3.4.3), we deduce

$$\left\| \delta v^\delta(x, p, \cdot) \right\|_{L^\infty} \leq \lambda(p). \quad (3.4.7)$$

We claim the following uniform (in δ) bound

$$\left\| v^\delta(x, p, \cdot) - v^\delta(x, p, 0) \right\|_{L^\infty} \leq C(1 + \lambda(p) + |p|). \quad (3.4.8)$$

If we assume that (3.4.8) does not hold, then there exists a sequence $(\delta_k, x_k, p_k)_{k \in \mathbb{N}}$ with

$\delta_k \rightarrow 0$ such that the solution $v_k(\cdot) = v^{\delta_k}(x_k, p_k, \cdot)$ satisfies

$$\|v_k - v_k(0)\|_{L^\infty} \geq k(1 + \lambda(p_k) + |p_k|).$$

We set $b_k = \|v_k - v_k(0)\|_{L^\infty}^{-1}$ and $\tilde{v}_k = b_k(v_k - v_k(0))$. Then, $\tilde{v}_k(0) = 0$, $\|\tilde{v}_k\|_{L^\infty} = 1$ and $\tilde{v}_k$ satisfies

$$\begin{aligned} \delta_k \tilde{v}_k + b_k \delta_k v_k(0) - \text{tr} \left(a D_{yy}^2 \tilde{v}_k \right) \\ + \sup_{\alpha \in \mathbb{R}^{2d}} \left\{ -b_k p_k \cdot c(x_k, y, \alpha) - D_y \tilde{v}_k \cdot f(x_k, y, \alpha) - b_k L(x_k, y, \alpha) \right\} = 0. \end{aligned}$$

Since L, c are Lipschitz with respect to y , we have for $\beta \in (0, 1)$ and due to (3.4.7)

$$|b_k \delta_k v_k(0)| + \|b_k p_k \cdot c(x_k, \cdot, \alpha)\|_{C^{0,\beta}} + \|b_k L(x_k, \cdot, \alpha)\|_{C^{0,\beta}} \leq C b_k (1 + |p_k| + \lambda(p_k)) \leq \frac{C}{k}.$$

Now, by the regularity theory of quasilinear elliptic equations [GT77, Saf89], we deduce that $\tilde{v}_k$ is uniformly bounded in $C^{2,\beta}$ for some $\beta \in (0, 1)$. Fix $\alpha_0 \in \mathcal{A}$. The families of functions $\{\tilde{v}_k(\cdot)\}$, $\{c(x_k, \cdot, \alpha_0)\}$, $\{f(x_k, \cdot, \alpha_0)\}$ are uniformly Lipschitz and equibounded, hence the Arzela-Ascoli theorem yields the existence of a limit (along subsequences) to $\tilde{v}, \tilde{c}, \tilde{f}$, respectively, where $\tilde{v} \in C^{2,\beta}$. It is clear that $\tilde{v}(0) = 0$, $\|\tilde{v}\|_{L^\infty} = 1$ and $\tilde{v}$ satisfies

$$-\text{tr} \left(a D_{yy}^2 \tilde{v} \right) - D_y \tilde{v} \cdot \tilde{f}(y) \leq 0. \quad (3.4.9)$$

Notice that $\tilde{v}$ is a periodic function, thus by the strong maximum principle in (3.4.9), it must be constant. This contradicts the previous equalities.

Now, we may use the estimates for quasilinear elliptic equations for the equation satisfied

by v^δ and assumptions (A1), (A2) to get

$$\begin{aligned}
\|v^\delta - v^\delta(0)\|_{C^{2,\beta}} &\leq C \left(\|v^\delta - v^\delta(0)\|_{L^\infty} + |\delta v^\delta(0)| \right. \\
&\quad \left. + \sup_{\alpha \in A} \|L(x, \cdot, \alpha)\|_{C^{0,\beta}} + \sup_{\alpha \in A} \|p \cdot c(x, \cdot, \alpha)\|_{C^{0,\beta}} \right) \\
&\stackrel{(3.4.8), (3.4.7)}{\leq} C(1 + |p| + \lambda(p)),
\end{aligned}$$

which is (3.4.6).

With (3.4.6) and (3.4.7) in mind, we can use Arzela-Ascoli once again to the uniformly bounded and equicontinuous family $\{v^\delta(x, p, \cdot) - v^\delta(x, p, 0)\}$ to extract a subsequence such that $v^\delta(x, p, y) - v^\delta(x, p, 0) \rightarrow v(x, p, y)$ in C^2 as $\delta \rightarrow 0$ and $\delta v^\delta(x, p, 0) \rightarrow -\bar{H}(x, p)$ (bounded sequence has convergent subsequence). Moreover, v is a classical solution of

$$-\text{tr} \left(a D_{yy}^2 v \right) + H(x, y, p, D_y v) + \bar{H}(x, p) = 0. \quad (3.4.10)$$

However, it is straightforward to show that there is a unique pair $(v, \bar{H}(x, p))$ satisfying (3.4.10), hence $\bar{H}(x, p), v$ are unique subsequential limits of $\delta v^\delta(0), v^\delta - v^\delta(0)$, respectively. Finally, (3.4.4) follows by the triangle inequality and the aforementioned convergences.

Step 2. (first derivatives of v^δ)

Throughout the proof we consider $(x, p) \in \mathbb{T}^d \times \mathbb{R}^d$ to be fixed.

We will start with the derivatives with respect to x . We will show that the partial derivatives $\partial_{x_i} v^\delta(x, p, y)$ exist for every $i = 1, \dots, d$. Let $i \in \{1, \dots, d\}$ and h be a vector which is parallel to $(0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^d$, where the 1 is in the i -th position. The difference $V_i^\delta(y) := v^\delta(x + h, p, y) - v^\delta(x, p, y)$ satisfies the equation

$$\delta V_i^\delta - \text{tr} \left(a D_{yy}^2 V_i^\delta \right) + \int_0^1 H_q \left(sh + x, y, p, s D_y V_i^\delta + D_y v^\delta(x, p, y) \right) ds \cdot D_y V_i^\delta$$

$$= -|h| \int_0^1 H_{x_i} \left(sh + x, y, p, sD_y V_i^\delta + D_y v^\delta(x, p, y) \right) ds. \quad (3.4.11)$$

Let $y_1, y_2 \in \mathbb{T}^d$ be a position of maximum and a position of minimum of V^δ , respectively.

By combining the optimality conditions with (3.4.11) we obtain

$$\begin{aligned} \frac{|h|}{\delta} \int_0^1 H_{x_i} \left(sh + x, y_2, p, D_y v^\delta(x, p, y_2) \right) ds &\leq \min_y V_i^\delta \\ \text{and } \max_y V_i^\delta &\leq \frac{|h|}{\delta} \int_0^1 H_{x_i} \left(sh + x, y_1, p, D_y v^\delta(x, p, y_1) \right) ds, \end{aligned}$$

hence, due to (3.4.6) and assumption (A3)(3) (equation (3.2.2)) on H ,

$$\|V_i^\delta(\cdot)\|_{L^\infty} \leq \frac{C_H}{\delta} (1 + |p|)|h| = C'|h|, \quad (3.4.12)$$

for some constant $C' > 0$. We also observe that, due to (3.4.6) and the assumptions on H , the two integrals in (3.4.11) are Lipschitz with respect to the y variable, therefore by the Schauder estimates for elliptic equations (see [GT77, Chapter 6]) and (3.4.12), there exists a constant $C > 0$ such that $\|V_i^\delta\|_{C^{2,\beta}} \leq C|h|$. By Arzela-Ascoli, this implies that $V_i^\delta/|h|$ has a convergent subsequence (in C^2) as $|h| \rightarrow 0$ to a function $w_i^\delta(x, p, y)$. By sending $|h| \rightarrow 0$ (along the subsequence) in (3.4.11), we get that the limit $w_i^\delta(x, p, \cdot) \in C^2$ satisfies the equation

$$\begin{aligned} \delta w_i^\delta - \text{tr} \left(a D_{yy}^2 w_i^\delta \right) \\ + H_q \left(x, y, p, D_y v^\delta(x, p, y) \right) \cdot D_y w_i^\delta = -H_{x_i} \left(x, y, p, D_y v^\delta(x, p, y) \right), \end{aligned} \quad (3.4.13)$$

in the classical sense. Since (3.4.13) has a unique solution, we deduce that $V_i^\delta(y)/|h| \xrightarrow{|h| \rightarrow 0} w_i^\delta(x, p, y)$ in C^2 . Thus, $\partial_{x_i} v^\delta(x, p, y) = w_i^\delta(x, p, y)$. We also note that $\partial_{y_j} \partial_{x_i} v^\delta(x, p, y) = \partial_{y_j} w_i^\delta(x, p, y)$ for $j = 1, \dots, d$. Furthermore, by the C^2 convergence we get $\partial_{x_i} \partial_{y_j} v^\delta = \partial_{y_j} w_i^\delta(x, p, y)$ for $j = 1, \dots, d$. By repeating the same argument in all directions, we conclude

that $D_x v^\delta(x, p, y)$ and $D_{yx}^2 v^\delta(x, y, p) = D_{xy}^2 v^\delta(x, p, y)$ exist and are continuous.

For the derivative with respect to p we use a similar strategy. Let $i \in \{1, \dots, d\}$ and h be a vector which is parallel to $(0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^d$, where 1 is in the i -th position, with $|h| < 1$. The difference $\tilde{V}_i^\delta(y) := v^\delta(x, p + h, y) - v^\delta(x, p, y)$ satisfies

$$\begin{aligned} & \delta \tilde{V}_i^\delta - \text{tr} \left(a D_{yy}^2 \tilde{V}_i^\delta \right) \\ & + \int_0^1 H_q \left(x, y, sh + p, s D_y \tilde{V}_i^\delta + D_y v^\delta(x, p, y) \right) ds \cdot D_y \tilde{V}_i^\delta \\ & = -|h| \int_0^1 H_{p_i} \left(x, y, sh + p, s D_y \tilde{V}_i^\delta + D_y v^\delta(x, p, y) \right) ds. \end{aligned} \quad (3.4.14)$$

By arguing as in the x -derivative case, by assumption (A3)(3) (equation (3.2.3)) on H we have

$$\|\tilde{V}_i^\delta(\cdot)\|_{L^\infty} \leq \frac{C_H}{\delta} (1 + |p| + |h|)|h| \leq C'|h|, \quad (3.4.15)$$

for some constant $C' > 0$, since $|h| < 1$ and p is fixed. In addition, the two integrals in (3.4.14) are β -Hölder continuous with respect to y , therefore by the Schauder estimates, we also have $\|\tilde{V}_i^\delta\|_{C^{2,\beta}} \leq C|h|$ for some constant $C > 0$. Now, as in the x -derivative case we can argue that $\tilde{V}_i^\delta/|h|$ converges in C^2 , as $|h| \rightarrow 0$, to the unique classical solution of

$$\begin{aligned} & \delta \tilde{w}_i^\delta - \text{tr} \left(a D_{yy}^2 \tilde{w}_i^\delta \right) + H_q \left(x, y, p, D_y v^\delta(x, p, y) \right) \cdot D_y \tilde{w}_i^\delta \\ & = -H_{p_i} \left(x, y, p, D_y v^\delta(x, p, y) \right). \end{aligned} \quad (3.4.16)$$

Thus, since there is a unique function satisfying (3.4.16), $\tilde{V}_i^\delta/|h| \xrightarrow{|h| \rightarrow 0} \tilde{e}_i^\delta$, which implies $\partial_{p_i} v^\delta(x, p, y) = \tilde{w}_i^\delta(x, p, y)$ and $\partial_{y_j} \partial_{p_1} v^\delta(x, p, y) = \partial_{y_j} \tilde{w}_i^\delta(x, p, y)$ for $j = 1, \dots, d$. Furthermore, by the convergence in C^2 , we also have $\partial_{p_i} \partial_{y_j} v^\delta(x, p, y) = \partial_{y_j} \tilde{w}_i^\delta(x, p, y)$ for $j = 1, \dots, d$. Therefore, we deduce that $D_p v^\delta, D_{py}^2 v^\delta, D_{yp}^2 v^\delta$ exist and $D_{yp}^2 v^\delta = D_{py}^2 v^\delta$.

Before continuing with the next step, we note that the bounds $\|V_i^\delta\|_{C^{2,\beta}} \leq C(1 + |p|)|h|$ and

$\|\tilde{V}_i^\delta\|_{C^{2,\beta}} \leq C(1+|p|)|h|$ for $i = 1, \dots, d$ imply that $x \mapsto D_y v^\delta(x, p, y)$ is Lipschitz continuous with Lipschitz constant $C(1+|p|)$ and the map $p \mapsto D_y v^\delta(x, p, y)$ is continuous.

Step 3. (second derivatives of v^δ)

We fix $i, j \in \{1, \dots, d\}$. We will first show that the map $x \mapsto D_y w_i^\delta(x, p, y)$ is continuous, where w_i^δ is the unique solution of (3.4.13). Let h be a vector parallel to $(0, \dots, 0, 1, 0, \dots, 0) \in \mathbb{R}^d$, where the 1 is in the j -th position. The function $W_{ij}^\delta(y) := w_i^\delta(x + h, p, y) - w_i^\delta(x, p, y)$ satisfies

$$\begin{aligned} & \delta W_{ij}^\delta - \text{tr} \left(a D_{yy}^2 W_{ij}^\delta \right) + H_q(x, y, p, D_y v^\delta) \cdot D_y W_{ij}^\delta \\ &= -|h| \int_0^1 H_{x_j x_i} \left(sh + x, y, p, s D_y V_j^\delta + D_y v^\delta \right) ds \\ & \quad - \int_0^1 \partial_{x_j} H_q \left(sh + x, y, p, s D_y V_j^\delta + D_y v^\delta \right) ds \cdot D_y V_j^\delta \\ & \quad - |h| \int_0^1 \partial_{x_j} H_q \left(sh + x, y, p, s D_y V_j^\delta + D_y v^\delta \right) ds \cdot D_y w_i^\delta(x + h) \\ & \quad - \int_0^1 H_{qq} \left(sh + x, y, p, s D_y V_j^\delta + D_y v^\delta \right) D_y V_j^\delta ds \cdot D_y w_i^\delta(x + h), \quad (3.4.17) \end{aligned}$$

where V_i^δ , for $i = 1, \dots, d$, was defined in (3.4.11). By Step 2, we have a uniform Hölder bound for w_i^δ , namely $\|w_i^\delta(x)\|_{C^{2,\beta}} \leq C(1+|p|)$ for some $C > 0$ independent of x . Combining this with (3.4.6) and $\|V^\delta\|_{C^{2,\beta}} \leq C|h|$ we have the following bounds for the terms on the right hand side of (3.4.17)

$$\begin{aligned} & \left| |h| \int_0^1 H_{x_j x_i} \left(sh + x, y, p, s D_y V_j^\delta + D_y v^\delta \right) ds \right| \leq C|h|, \\ & \left| \int_0^1 \partial_{x_j} H_q \left(sh + x, y, p, s D_y V_j^\delta + D_y v^\delta \right) ds \cdot D_y V_j^\delta \right| \leq C|h|, \\ & \left| |h| \int_0^1 \partial_{x_j} H_q \left(sh + x, y, p, s D_y V_j^\delta + D_y v^\delta \right) ds \cdot D_y w_i^\delta(x + h) \right| \leq C|h|, \\ & \left| \int_0^1 H_{qq} \left(sh + x, y, p, s D_y V_j^\delta + D_y v^\delta \right) D_y V_j^\delta ds \cdot D_y w_i^\delta(x + h) \right| \leq C|h|. \end{aligned}$$

Thus, by the maximum principle, we have $\|W_{ij}^\delta(\cdot)\|_{L^\infty} \leq C|h|$. In addition, by the properties of H , all these terms are Lipschitz continuous in the y variable, hence by the Schauder estimates we get $\|W_{ij}^\delta(\cdot)\|_{C^{2,\beta}} \leq C|h|$. Since j was arbitrary, these estimates yield the continuity of $x \mapsto D_y w_i^\delta(x, p, y)$.

We introduce the function $\widetilde{W}_{ij}^\delta(y) = \widetilde{w}_i^\delta(x, p+h, y) - \widetilde{w}_i^\delta(x, p, y)$, where i, j, h are the same as above and $\widetilde{w}_i^\delta$ is the solution of (3.4.16). We will show that the map $p \mapsto D_y \widetilde{w}_i^\delta(x, p, y)$ is continuous. Indeed, $\widetilde{W}_{ij}^\delta$ satisfies

$$\begin{aligned} & \delta \widetilde{W}_{ij}^\delta - \text{tr} \left(a D_{yy}^2 \widetilde{W}_{ij}^\delta \right) + H_q(x, y, p, D_y v^\delta(p)) \cdot D_y \widetilde{W}_{ij}^\delta \\ &= -|h| \int_0^1 H_{p_j p_i} \left(x, y, sh + p, s D_y \widetilde{V}_j^\delta + D_y v^\delta \right) ds \\ & \quad - \int_0^1 \partial_{p_j} H_q \left(x, y, sh + p, s D_y \widetilde{V}_j^\delta + D_y v^\delta \right) ds \cdot D_y \widetilde{V}_j^\delta \\ & \quad - |h| \int_0^1 \partial_{p_j} H_q \left(x, y, sh + p, s D_y \widetilde{V}_j^\delta + D_y v^\delta \right) ds \cdot D_y \widetilde{w}_i^\delta(p+h) \\ & \quad - \int_0^1 H_{qq} \left(x, y, sh + p, s D_y \widetilde{V}_j^\delta + D_y v^\delta \right) D_y \widetilde{V}_j^\delta ds \cdot D_y \widetilde{w}_i^\delta(p+h), \end{aligned} \quad (3.4.18)$$

where $\widetilde{V}_j^\delta$, for $j = 1, \dots, d$ was defined in (3.4.14). Again by Step 2, we have the bounds $\|\widetilde{w}_i^\delta\|_{C^{2,\beta}} \leq C$ and $\|\widetilde{V}_i^\delta\|_{C^{2,\beta}} \leq C|h|$, where C depends on p . Thus, we can show with a similar argument as before this time for equation (3.4.18), that the map $p \mapsto D_y \widetilde{w}_i^\delta(x, p, y)$ is continuous.

Using the continuity of both maps $x \mapsto D_y w_i^\delta(x, p, y)$ and $p \mapsto D_y \widetilde{w}_i^\delta(x, p, y)$, we can prove that $D_{xx}^2 v^\delta, D_{xp}^2 v^\delta, D_{px}^2 v^\delta$ and $D_{xx}^2 v^\delta$ exist and are continuous. Indeed, in light of these properties, the right hand side in (3.4.13) and the coefficient of $D_y w_i^\delta$ in (3.4.13) are C^1 functions with respect to the x and p variables, therefore we can apply Lemma 3.6.4 for (3.4.13) to deduce that $(x, p) \mapsto w_i^\delta(x, p, y) = \partial_{x_i} v^\delta$ is also a C^1 function, $i = 1, \dots, d$. Hence, $D_{xx}^2 v^\delta$ and $D_{px}^2 v^\delta$ exist and are continuous. We can similarly apply Lemma 3.6.4 to (3.4.16) to deduce that $D_{pp}^2 v^\delta$ and $D_{xp}^2 v^\delta$ exist and are continuous.

Step 4. (regularity of $\overline{H}$, (3.4.5))

Let $x, x' \in \mathbb{T}^d$ and $p, p' \in \mathbb{R}^d$. By (3.4.12) for $i = 1, \dots, d$, we get for any $y \in \mathbb{T}^d$

$$|\delta v^\delta(x', p', y) - \delta v^\delta(x, p', y)| \leq |x' - x|(1 + |p'|). \quad (3.4.19)$$

On the other hand, by (3.4.15) for $i = 1, \dots, d$ (note that the condition $|h| < 1$ is used only on the second part of the inequality) we have

$$|\delta v^\delta(x, p', y) - \delta v^\delta(x, p, y)| \leq C_H |p - p'|(|p| + |p'|) \quad (3.4.20)$$

We add (3.4.19), (3.4.20) and we send $\delta \rightarrow 0$ to discover, by virtue of (3.4.4),

$$|\overline{H}(x', p') - \overline{H}(x, p')| + |\overline{H}(x, p') - \overline{H}(x, p)| \leq C_H(1 + |p| + |p'|)(|x - x'| + |p - p'|).$$

(3.4.5) follows by applying the triangle inequality in the above.

□

We are now ready to prove Theorem 3.1.1.

Proof. Let $\overline{H}$ be the function we constructed in Proposition 3.4.1. By Corollary 3.3.7, the family $U^\varepsilon(t, m)$ is uniformly bounded (the uniform bound depends only on $\mathcal{F}, \mathcal{G}$ and $\|H(\cdot, \cdot, 0)\|_{L^\infty}$), hence we can define the upper and lower semi-limits for $(t, \mu) \in [0, T] \times \mathcal{P}(\mathbb{T}^d)$:

$$\overline{U}(t, \mu) = \limsup_{\substack{\varepsilon \rightarrow 0 \\ t' \rightarrow t \\ \mu' \rightarrow \mu}} \sup_{\substack{m \in \mathcal{P}(\mathbb{T}^{2d}): \\ \pi_*^1 m = \mu'}} U^\varepsilon(t', m), \quad (3.4.21)$$

$$\underline{U}(t, \mu) = \liminf_{\substack{\varepsilon \rightarrow 0 \\ t' \rightarrow t \\ \mu' \rightarrow \mu}} \inf_{\substack{m \in \mathcal{P}(\mathbb{T}^{2d}): \\ \pi_*^1 m = \mu'}} U^\varepsilon(t', m). \quad (3.4.22)$$

We will show that $\bar{U}$ is a viscosity subsolution of (3.1.4) and that $\underline{U}$ is a viscosity supersolution of (3.1.4). Then, since (3.4.5) holds, Theorem 3.3.6 applies for (3.1.4); therefore $\underline{U} \geq \bar{U}$. Combining this with the obvious fact that $\underline{U} \leq \bar{U}$, we deduce $\underline{U} = \bar{U} := U$. Since both semi-limits are equal to U , the uniform convergence of U^ε to U that we need to prove. We will only show that $\bar{U}$ is a subsolution, the proof for the supersolution being analogous.

To prove the first part of the definition $\bar{U}(T, z, \mu) \leq \tilde{\mathcal{G}}(z, \mu)$ for any $(z, \mu) \in \mathbb{T}^d \times \mathcal{P}(\mathbb{T}^d)$, we rely on [CM23, Theorem 4.2]⁵ to get a sequence $(\mathcal{G}^n)_{n \in \mathbb{N}} : \mathcal{P}(\mathbb{T}^d) \rightarrow \mathbb{R}$ of uniformly Lipschitz and $C^{1,2}$ functions such that $\mathcal{G}^n \xrightarrow{n \rightarrow \infty} \mathcal{G}$ uniformly. For any $\delta > 0$ there exists $n_0 \in \mathbb{N}$ such that $\mathcal{G}(\mu) \leq \mathcal{G}^n(\mu) + \delta$ for any $n \geq n_0$ and $\mu \in \mathcal{P}(\mathbb{T}^d)$. The regularity of $\mathcal{G}^n$ implies that there exists a constant $C_n > 0$ such that $V^{n,\delta}(t, m) := \mathcal{G}^n(\pi_*^1 m) + C_n(T - t) + \delta$ is a classical supersolution of (3.1.1), where $n \geq n_0$. Therefore, by comparison principle, we have $U^\varepsilon(t, m) \leq V^{n,\delta}(t, m)$. By taking the upper semi-limit in this inequality we get $\bar{U}(T, \mu) \leq \mathcal{G}^n(\mu) + \delta$ for any $n \geq n_0$. The result follows by sending $n \rightarrow \infty$ and $\delta \rightarrow 0$.

For the second part we argue by contradiction. Assume that there exists a test function (Definitions 3.3.1, 3.3.3 and 3.3.5) φ and a point $(\bar{t}, \bar{z}, \bar{\mu}) \in [0, T) \times \mathbb{T}^d \times \mathcal{P}(\mathbb{T}^d)$ such that $\bar{U}(t, z, \mu) - \varphi(t, z, \mu)$ has a strict maximum with value 0 (recall Remark 3.3.4) at $(\bar{t}, \bar{z}, \bar{\mu})$ and

$$\begin{aligned} -\partial_t \varphi(\bar{t}, \bar{z}, \bar{\mu}) + \int_{\mathbb{T}^d} \bar{H}(x + \bar{z}, D_\mu \varphi(\bar{t}, \bar{z}, \bar{\mu}, x)) \bar{\mu}(dx) - \mathcal{F}(\bar{\mu}) \\ - \int_{\mathbb{T}^d} \text{tr} \left(\sigma_1 \sigma_1^\top D_x D_\mu \varphi(\bar{t}, \bar{\mu}, \bar{\mu}, x) \right) \bar{\mu}(dx) \\ - \text{tr} \left(\sigma_{0,1} \sigma_{0,1}^\top D_{zz}^2 \varphi(\bar{t}, \bar{z}, \bar{\mu}) \right) \geq 10\eta. \end{aligned} \quad (3.4.23)$$

for some $\eta > 0$. We will assume that $\bar{t} > 0$; the case $\bar{t} = 0$ can be treated similarly. Note also that, due to Lemma 3.6.5 and (3.4.5), we may assume that φ is such that $D_{xx}^2 D_m \varphi$ and $D_x D_m \varphi$ exist and are continuous with (3.4.23) still holding (perhaps for a different η)

5. The result in this paper is stated for functions defined on $\mathcal{P}_2(\mathbb{R}^d)$, but the proof can be easily adapted for functions defined on $\mathcal{P}(\mathbb{T}^d)$.

and $\widetilde{U}(t, z, \mu) - \varphi(t, z, \mu)$ having a strict maximum with value 0 at $(\bar{t}, \bar{z}, \bar{\mu})$ (perhaps for a different $(\bar{t}, \bar{z}, \bar{\mu})$). By Proposition 3.4.1 after setting $x \mapsto x + \bar{z}$ and $p \mapsto D_\mu \varphi(\bar{t}, \bar{z}, \bar{\mu}, x)$, for every $\delta > 0$ and $x, z \in \mathbb{T}^d$ there exists a function $v_{x,z}^\delta$ such that

$$\begin{aligned} \delta v^\delta - \text{tr} \left((\sigma_2 \sigma_2^\top + \sigma_{0,2} \sigma_{0,2}^\top) D_{yy}^2 v^\delta \right) \\ + H \left(x + \bar{z}, y + z, D_\mu \varphi(\bar{t}, \bar{z}, \bar{\mu}, x), D_y v^\delta \right) = 0, \quad y \in \mathbb{T}^d. \end{aligned} \quad (3.4.24)$$

It is clear that v^δ has only two arguments $v_{x,z}^\delta(\cdot) = v^\delta(x, \cdot + z)$ and, furthermore, δ can be chosen small enough such that

$$\left\| \delta v^\delta(x, z + \cdot) + \overline{H}(x + \bar{z}, D_\mu \varphi(\bar{t}, \bar{z}, \bar{\mu}, x)) \right\|_{L^\infty} \leq \eta, \quad \text{for any } x, z \in \mathbb{T}^d. \quad (3.4.25)$$

We now consider the function $\psi^{\varepsilon, \delta} : [0, T] \times \mathbb{T}^d \times \mathbb{T}^d \times \mathcal{P}(\mathbb{T}^{2d})$ with $\psi^{\varepsilon, \delta}(t, z_1, z_2, m) := \varphi(t, z_1, \pi_*^1 m) + \varepsilon \int_{\mathbb{T}^{2d}} v^\delta(x, y + z_2) m(dx, dy)$. For $\mu \in \mathcal{P}(\mathbb{T}^d)$, we have by the uniform convergence of $\psi^{\varepsilon, \delta}$ to φ (as $\varepsilon \rightarrow 0$)

$$\begin{aligned} \limsup_{\substack{\varepsilon \rightarrow 0 \\ t' \rightarrow t \\ \mu' \rightarrow \mu}} \sup_{\substack{m \in \mathcal{P}(\mathbb{T}^{2d}): \\ \pi_*^1 m = \mu'}} (\widetilde{U}^\varepsilon - \psi^{\varepsilon, \delta})(t', z_1, z_2, m) &= \limsup_{\substack{\varepsilon \rightarrow 0 \\ t' \rightarrow t \\ \mu' \rightarrow \mu}} \sup_{\substack{m \in \mathcal{P}(\mathbb{T}^{2d}): \\ \pi_*^1 m = \mu'}} \widetilde{U}^\varepsilon(t', z_1, z_2, m) - \varphi(t, z_1, \mu) \\ &= \limsup_{\substack{\varepsilon \rightarrow 0 \\ t' \rightarrow t \\ \mu' \rightarrow (\text{Id} + z_1)_* \mu}} \sup_{\substack{m \in \mathcal{P}(\mathbb{T}^{2d}): \\ \pi_*^1 m = \mu'}} U^\varepsilon(t', m) - \varphi(t, z_1, \mu) \\ &= \overline{U}(t, (\text{Id} + z_1)_* \mu) - \varphi(t, z_1, \mu) \\ &= \widetilde{U}(t, z_1, \mu) - \varphi(t, z_1, \mu). \end{aligned}$$

Since $(\bar{t}, \bar{z}, \bar{\mu})$ is a position of strict maximum for $\widetilde{U} - \varphi$, there exists an $r > 0$ small enough such that the above limit is negative for every $(t, z_1, \mu) \in (\bar{t} - r, \bar{t} + r) \times \{z : \mathbf{d}_{\mathbb{T}^d}(z, \bar{z}) \leq r\} \times B_r(\mu)$, where $B_r(\mu) = \{\mu \in \mathcal{P}(\mathbb{T}^d) : \mathbf{d}_1(\mu, \bar{\mu}) \leq r\}$ and $\mathbf{d}_{\mathbb{T}^d}$ is the distance between points in $\mathbb{T}^d$. However, by compactness, this implies that there exists an $\tilde{\eta} > 0$ such that

$\tilde{U}^\varepsilon(t, z_1, z_2, m) - \psi^{\varepsilon, \delta}(t, z_1, z_2, m) \leq -\tilde{\eta}$ for every (t, z_1, z_2, m) on the boundary of the set

$$A_r := (\bar{t} - r, \bar{t} + r) \times \{(z_1, z_2) \in \mathbb{T}^{2d} : \mathbf{d}_{\mathbb{T}^d}(z_1, \bar{z}) \leq r\} \times \left\{m \in \mathcal{P}(\mathbb{T}^d) : \mathbf{d}_1(\pi_*^1 m, \bar{\mu}) < r\right\}.$$

We claim that for ε small enough

$$\tilde{U}^\varepsilon(t, z_1, z_2, m) - \psi^{\varepsilon, \delta}(t, z_1, z_2, m) \leq -\tilde{\eta}, \quad \text{in } A_r. \quad (3.4.26)$$

Before proving (3.4.26), we explain why it gives a contradiction. Indeed, if (3.4.26) holds, then taking the semi-limit on both sides yields

$$\tilde{U}(t, z_1, \mu) - \varphi(t, z_1, \mu) \leq -\tilde{\eta},$$

for every $t \in (\bar{t} - r, \bar{t} + r)$, $z_1 \in \{z \in \mathbb{T}^d : \mathbf{d}_{\mathbb{T}^d}(z, \bar{z}) \leq r\}$ and $\mu \in B_r$. In particular, setting $t = \bar{t}$, $z_1 = \bar{z}$ and $\mu = \bar{\mu}$, we get a contradiction.

It remains to show (3.4.26). If (3.4.26) is not true, we may suppose that

$$\max_{A_r} \left\{ \tilde{U}^\varepsilon(t, z_1, z_2, m) - \psi^{\varepsilon, \delta}(t, z_1, z_2, m) \right\} = \tilde{U}^\varepsilon(t_0, z_{0,1}, z_{0,2}, m_0) - \psi^{\varepsilon, \delta}(t_0, z_{0,1}, z_{0,2}, m_0),$$

where $(t_0, z_{0,1}, z_{0,2}, m_0)$ is an interior point of A_r . We observe that, due to the regularity of v^δ proved in Proposition 3.4.1, $\psi^{\varepsilon, \delta}$ is a test function for (3.1.1) (see Definition 3.3.1), hence by the definition of viscosity solutions

$$\begin{aligned} 0 &\geq -\partial_t \varphi(t_0, z_{0,1}, \pi_*^1 m_0) - \mathcal{F}(\pi_*^1 m_0) \\ &\quad + \int_{\mathbb{T}^{2d}} H\left(x + z_{0,1}, y + z_{0,2}, D_\mu \varphi(t_0, z_{0,1}, \pi_*^1 m_0, x) + \varepsilon D_x v^\delta(x, y + z_{0,2}) \right. \\ &\quad \left. , D_y v^\delta(x, y + z_{0,2})\right) m_0(dx, dy) \end{aligned}$$

$$\begin{aligned}
& - \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_1 \sigma_1^\top D_x D_\mu \varphi(t_0, z_{0,1}, \pi_*^1 m_0, x) \right) \pi_*^1 m_0(dx) \\
& - \varepsilon \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_1 \sigma_1^\top D_{xx}^2 v^\delta(x, y + z_{0,2}) \right) m_0(dx, dy) \\
& - \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_2 \sigma_2^\top D_{yy}^2 v^\delta(x, y + z_{0,2}) \right) m_0(dx, dy) \\
& - \text{tr} \left(\sigma_{0,1} \sigma_{0,1}^\top D_{z_1 z_1}^2 \varphi(t_0, z_{0,1}, \pi_*^1 m_0) \right) \\
& - \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_{0,2} \sigma_{0,2}^\top D_{z_2 z_2}^2 v^\delta(x, y + z_{0,2}) \right) m_0(dx, dy). \tag{3.4.27}
\end{aligned}$$

Since r is small enough and $(t_0, z_{0,1}, z_{0,2}, m_0) \in A_r$, due to the uniform continuity of $\partial_t \varphi$, $D_{zz}^2 \varphi$ and $D_x D_\mu \varphi$ and the Lipschitz continuity of $\mathcal{F}$ and $D_x D_\mu \varphi(t, z_1, \mu, \cdot)$, we may assume that

$$|\partial_t \varphi(t_0, z_{0,1}, \pi_*^1 m_0) - \partial_t \varphi(\bar{t}, \bar{z}, \bar{\mu})| \leq \tilde{C} \eta, \tag{3.4.28}$$

$$\begin{aligned}
& \left| \int_{\mathbb{T}^d} \text{tr} \left(\sigma_1 \sigma_1^\top D_x D_\mu \varphi(t_0, z_{0,1}, \pi_*^1 m_0, x) \right) \pi_*^1 m_0(dx) \right. \\
& \quad \left. - \int_{\mathbb{T}^d} \text{tr} \left(\sigma_1 \sigma_1^\top D_x D_\mu \varphi(\bar{t}, \bar{z}, \bar{\mu}, x) \right) \bar{\mu}(dx) \right| \leq \tilde{C} \eta, \tag{3.4.29}
\end{aligned}$$

$$\left| \text{tr} \left(\sigma_{0,1} \sigma_{0,1}^\top D_{zz}^2 \varphi(t_0, z_{0,1}, \pi_*^1 m_0) \right) - \text{tr} \left(\sigma_{0,1} \sigma_{0,1}^\top D_{zz}^2 \varphi(\bar{t}, \bar{z}, \bar{\mu}) \right) \right| \leq \tilde{C} \eta \tag{3.4.30}$$

and

$$|\mathcal{F}(\pi_*^1 m_0) - \mathcal{F}(\bar{\mu})| \leq \text{Lip}(\mathcal{F}) \mathbf{d}_1(\pi_*^1 m_0, \bar{\mu}) \leq \text{Lip}(\mathcal{F}) r \leq \tilde{C} \eta, \tag{3.4.31}$$

for some constant $\tilde{C}$ small enough. In addition, by the properties of H (assumption (A3), equation (3.2.1)) and the uniform continuity of $D_\mu \varphi$ we have

$$\begin{aligned}
& \left| H \left(x + z_{0,1}, y + z_{0,2}, D_\mu \varphi(t_0, z_{0,1}, \pi_*^1 m_0, x) + \varepsilon D_x v^\delta(x, y + z_{0,2}), D_y v^\delta(x, y + z_{0,2}) \right) \right. \\
& \quad \left. - H \left(x + \bar{z}, y + z_{0,2}, D_\mu \varphi(\bar{t}, \bar{z}, \bar{\mu}, x), D_y v^\delta(x, y + z_{0,2}) \right) \right| \\
& \leq C_H \left(1 + \|D_y v^\delta\|_\infty + \varepsilon \|D_x v^\delta\|_\infty + \|D_\mu \varphi\|_\infty \right) \\
& \quad \cdot \left(|D_\mu \varphi(t_0, z_{0,1}, \pi_*^1 m_0, x) - D_\mu \varphi(\bar{t}, \bar{z}, \bar{\mu}, x)| + |z_{0,1} - \bar{z}| + \varepsilon \|D_x v^\delta\|_\infty \right)
\end{aligned}$$

$$\leq C \left(|D_\mu \varphi(t_0, z_{0,1}, \pi_*^1 m_0, x) - D_\mu \varphi(\bar{t}, \bar{z}, \bar{\mu}, x)| + r + \varepsilon \right) \leq \eta + C\varepsilon, \quad (3.4.32)$$

for some constant C and r small enough. We combine (3.4.28), (3.4.29), (3.4.30), (3.4.31), (3.4.32) with (3.4.27) to discover

$$\begin{aligned} & -\partial_t \varphi(\bar{t}, \bar{z}, \bar{\mu}) + \int_{\mathbb{T}^{2d}} H \left(x + \bar{z}, y + z_{0,2}, D_\mu \varphi(\bar{t}, \bar{z}, \bar{\mu}, x), D_y v^\delta(x, y + z_{0,2}) \right) m_0(dx, dy) - \mathcal{F}(\bar{\mu}) \\ & - \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_1 \sigma_1^\top D_x D_\mu \varphi(\bar{t}, \bar{z}, \bar{\mu}, x) \right) \bar{\mu}(dx) \\ & - \varepsilon \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_1 \sigma_1^\top D_{xx}^2 v^\delta(x, y + z_{0,2}) \right) m_0(dx, dy) \\ & - \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_2 \sigma_2^\top D_{yy}^2 v^\delta(x, y + z_{0,2}) \right) m_0(dx, dy) \\ & - \text{tr} \left(\sigma_{0,1} \sigma_{0,1}^\top D_{zz}^2 \varphi(\bar{t}, \bar{z}, \bar{\mu}) \right) - \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_{0,2} \sigma_{0,2}^\top D_{z_2 z_2}^2 v^\delta(x, y + z_{0,2}) \right) m_0(dx, dy) \\ & \leq 4\eta + C\varepsilon. \end{aligned} \quad (3.4.33)$$

Now, using the fact that $D_{zz}^2 v^\delta(x, y + z) = D_{yy}^2 v^\delta(x, y + z)$ combined with (3.4.24) and the continuity of $D_{xx}^2 v^\delta$ provided by Proposition 3.4.1, (3.4.33) becomes

$$\begin{aligned} & -\partial_t \varphi(\bar{t}, \bar{z}, \bar{\mu}) - \int_{\mathbb{T}^{2d}} \delta v^\delta(x, y + z_{0,2}) m_0(dx, dy) - \mathcal{F}(\bar{\mu}) - \varepsilon C_1 \\ & - \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_1 \sigma_1^\top D_x D_\mu \varphi(\bar{t}, \bar{z}, \bar{\mu}, x) \right) \bar{\mu}(dx) \\ & - \text{tr} \left(\sigma_{0,1} \sigma_{0,1}^\top D_{zz}^2 \varphi(\bar{t}, \bar{z}, \bar{\mu}) \right) \leq 4\eta + C\varepsilon, \end{aligned} \quad (3.4.34)$$

for some constant C_1 . However, with the help of estimate (3.4.25), this implies

$$\begin{aligned} & -\partial_t \varphi(\bar{t}, \bar{z}, \bar{\mu}) + \int_{\mathbb{T}^d} \bar{H}(x + \bar{z}, D_\mu \varphi(\bar{t}, \bar{z}, \bar{\mu}, x)) \pi_*^1 m_0(dx) - \mathcal{F}(\bar{\mu}) \\ & - \int_{\mathbb{T}^{2d}} \text{tr} \left(\sigma_1 \sigma_1^\top D_x D_\mu \varphi(\bar{t}, \bar{\mu}, x) \right) \bar{\mu}(dx) \\ & - \text{tr} \left(\sigma_{0,1} \sigma_{0,1}^\top D_{zz}^2 \varphi(\bar{t}, \bar{z}, \bar{\mu}) \right) \leq 3\eta + \varepsilon(C_1 + C). \end{aligned} \quad (3.4.35)$$

Finally, due to (3.4.5) and the regularity φ , $\overline{H}(x + \bar{z}, D_\mu \varphi(\bar{t}, \bar{z}, \bar{\mu}, x))$ is Lipschitz in x , so for r small enough we can write

$$\left| \int_{\mathbb{T}^d} \overline{H}(x + \bar{z}, D_\mu \varphi(\bar{t}, \bar{z}, \bar{\mu}, x)) (\pi_*^1 m_0 - \bar{\mu})(dx) \right| \leq C' \mathbf{d}_1(\pi_*^1 m_0, \bar{\mu}) \leq C' r \leq \eta.$$

Since C, C_1 , depend only on $\mathcal{F}, \varphi$ and the C^2 -norm of v^δ (which is independent of δ , by our arguments in Proposition (3.4.1)) we may also assume that ε is small enough such that $\varepsilon(C + C_1) \leq \eta$, hence (3.4.35) becomes

$$\begin{aligned} & -\partial_t \varphi(\bar{t}, \bar{z}, \bar{\mu}) + \int_{\mathbb{T}^d} \overline{H}(x + \bar{z}, D_\mu \varphi(\bar{t}, \bar{z}, \bar{\mu}, x)) \bar{\mu}(dx) - \mathcal{F}(\bar{\mu}) \\ & - \int_{\mathbb{T}^d} \text{tr} \left(\sigma_1 \sigma_1^\top D_x D_\mu \varphi(\bar{t}, \bar{\mu}, \bar{\mu}, x) \right) \bar{\mu}(dx) - \text{tr} \left(\sigma_{0,1} \sigma_{0,1}^\top D_{zz}^2 \varphi(\bar{t}, \bar{z}, \bar{\mu}) \right) \leq 4\eta. \end{aligned}$$

This contradicts (3.4.23). □

We finally give a proof of Corollary 3.1.3.

Proof. Under our assumptions, $U^\varepsilon(t, m) = \mathcal{G} \left(\mathcal{L}(X_T^{\varepsilon, t, m}) \right)$ is the unique viscosity solution of (3.1.1). In addition, for $x \in \mathbb{T}^d, p \in \mathbb{R}^d$, the effective Hamiltonian $\overline{H}(x, p)$ derived in Proposition 3.4.1 is such that the following equation has a unique solution

$$-\text{tr} \left((\sigma_2 \sigma_2^\top + \sigma_{0,2} \sigma_{0,2}^\top) D_{yy}^2 v \right) - p \cdot c(x, y) - f(x, y) \cdot D_y v - \overline{H}(x, p) = 0, \quad y \in \mathbb{T}^d. \quad (3.4.36)$$

It was proved in [BLP11] (see also [BS23a, Zit23]) that $\overline{H}(x, p) = p \cdot \int_{\mathbb{T}^d} c(x, y) \pi_x(dy)$, where π_x is the invariant measure associated to the problem $-\text{tr}(\sigma_2 \sigma_2^\top D_{yy}^2 v) - f(x, y) \cdot D_y v = 0$ and the integration is meant to be coordinate by coordinate. We set $\bar{c}(x) := \int_{\mathbb{T}^d} c(x, y) \pi_x(dy)$. By Theorem 3.1.1 we have that the limit of U^ε is the unique viscosity solution of the linear

Hamilton-Jacobi equation posed in $[t_0, T] \times \mathcal{P}(\mathbb{T}^d)$

$$\begin{cases} -\partial_t U + \int_{\mathbb{T}^d} \bar{c}(x) \cdot D_\mu U(t, \mu, x) \mu(dx) - \int_{\mathbb{T}^d} \text{tr} \left((\sigma_1 \sigma_1^\top + \sigma_{0,1} \sigma_{0,1}^\top) D_x D_\mu U(t, \mu, x) \right) \mu(dx) \\ \quad - \int_{\mathbb{T}^d} \int_{\mathbb{T}^d} \text{tr} \left(\sigma_{0,1} \sigma_{0,1}^\top D_{\mu\mu}^2 U(t, \mu, x, x') \right) \mu(dx) \mu(dx') = \mathcal{F}(\mu), \\ U(T, \mu) = \mathcal{G}(\mu). \end{cases}$$

This is the Hamilton-Jacobi equation associated with (3.1.8), hence following a standard dynamic programming argument⁶, we deduce that $U(t, m) = \mathcal{G} \left(\mathcal{L}(X_T^{t, \pi_*^1 m}) \right)$ is its viscosity solution. The proof is complete. $\square$

3.5 Wasserstein space and Wasserstein Derivatives

In this section, for the reader's convenience, we will gather the definitions of Wasserstein spaces and Wasserstein derivatives as well as some remarks which play an important role in the proof of our results.

Definition 3.5.1. Let $n_0 \in \mathbb{N}$ and $p \geq 1$. We define the p -Wasserstein space as

$$\mathcal{P}_p(\mathbb{T}^{n_0}) = \left\{ \mu \text{ probability measure over } \mathbb{T}^{n_0} \mid M_p(\mu) := \int_{\mathbb{T}^{n_0}} |x|^p \mu(dx) < +\infty \right\}$$

with metric $\mathbf{d}_p(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \left(\int_{\mathbb{T}^{2n_0}} |x - y|^p \pi(dxdy) \right)^{1/p}$. For $p = 1$ we also have $\mathbf{d}_1(\mu, \nu) = \sup_{f \in 1\text{-Lip}} \int_{\mathbb{T}^{n_0}} f(x) (\mu - \nu)(dx)$.

Remark 3.5.2. (i) It is known that for any $p \geq 1$, the metrics $\mathbf{d}_p$ are equivalent.

(ii) The p -Wasserstein space $\mathcal{P}_p(\mathbb{T}^{n_0})$ is compact.

(iii) Let $p > 1$. There exists $C := C(p) > 0$ such that for any $\mu_1, \mu_2 \in \mathcal{P}_p(\mathbb{T}^{n_0})$ we have $\mathbf{d}_p(\mu_1, \mu_2) \geq \mathbf{d}_1(\mu_1, \mu_2) \geq C \mathbf{d}_p(\mu_1, \mu_2)^p$.

⁶ See Proposition 3.6.1 for the argument without common noise ($\sigma_{0,1} = 0$) and in the case where there is control.

Lemma 3.5.3. *For any $M > 0$, the set $A_M = \{m \in \mathcal{P}(\mathbb{T}^{n_0}) | M_2(m) \leq M\}$ is compact in $\mathcal{P}_2(\mathbb{T}^{n_0})$.*

Proof. Since $\mathcal{P}_2(\mathbb{T}^{n_0})$ is compact, it suffices to show that A_M is closed. Indeed, let $(m_n)_{n \in \mathbb{N}}$ be a sequence in A_M such that $m_n \xrightarrow{\mathbf{d}_2} m$. For every $N \in \mathbb{N}$ we have

$$M \geq M_2(m_n) \geq \int_{\mathbb{T}^{n_0}} \min\{|x|^2, N\} m_n(dx).$$

We let $n \rightarrow +\infty$ and by the $\mathbf{d}_2$ convergence, since $\min\{|x|^2, N\}$ is bounded and continuous, we get $M \geq \int_{\mathbb{T}^{n_0}} \min\{|x|^2, N\} m(dx)$. We send N to infinity and by the monotone convergence theorem we deduce $M \geq M_2(m)$, hence $m \in A_M$. The proof is complete. $\square$

Definition 3.5.4. Let $U : \mathcal{P}_p(\mathbb{T}^{n_0}) \rightarrow \mathbb{R}$ be a function for some $p \geq 1$. We say that U is continuously differentiable or that it is C^1 in $\mathcal{P}_p(\mathbb{T}^{n_0})$ if there exists a continuous function $\frac{\delta U}{\delta m} : \mathcal{P}_p(\mathbb{T}^{n_0}) \times \mathbb{T}^{n_0} \rightarrow \mathbb{R}$ such that for every $m_1, m_2 \in \mathcal{P}_p(\mathbb{T}^{n_0})$ we have

$$\lim_{t \rightarrow 0} \frac{U(tm_1 + (1-t)m_2) - U(m_2)}{t} = \int_{\mathbb{T}^{n_0}} \frac{\delta U}{\delta m}(m_2, x)(m_2 - m_2)(dx).$$

If the x -derivative exists, we use the notation $D_m U(m, x) := D_x \frac{\delta U}{\delta m}(m, x)$, for $x \in \mathbb{T}^{n_0}$ and $m \in \mathcal{P}_p(\mathbb{T}^{n_0})$.

We finish the section with the following proposition, which is taken from [Zit23, Proposition A.5].

Proposition 3.5.5. *Let $U : \mathcal{P}_p(\mathbb{T}^{n_0}) \rightarrow \mathbb{R}$ be a C^1 function and $\pi^i : \mathbb{T}^{2n_0} \rightarrow \mathbb{T}^{n_0}$ such that $\pi^i(x_1, x_2) = x_i$, $i = 1, 2$. Then, for $i = 1, 2$ the function $\tilde{U}_i : \mathcal{P}_p(\mathbb{T}^{2n_0}) \rightarrow \mathbb{R}$ with $\tilde{U}_i(m) = U(\pi_*^i m)$ is also C^1 with*

$$\frac{\delta \tilde{U}_i}{\delta m}(m, x_1, x_2) = \frac{\delta U}{\delta \mu}(\pi_*^i m, \pi^i(x_1, x_2)), \text{ for every } m \in \mathcal{P}_p(\mathbb{T}^{2n_0}), x_1, x_2 \in \mathbb{T}^{n_0}. \quad (3.5.1)$$

In particular, $D_m \tilde{U}_i(m, x_1, x_2) = D_\mu U(\pi_*^i m, x_i)$ for every $m \in \mathcal{P}_p(\mathbb{T}^{2n_0})$ and $x_1, x_2 \in \mathbb{T}^{n_0}$, $i = 1, 2$.

Proof. We will show the result for $i = 1$ as the other case can be shown with an identical argument. For simplicity in the calculations below we drop the index i , so $\tilde{U}_i = \tilde{U}$ and $\pi^1 = \pi$. Let $m_1, m_2 \in \mathcal{P}_p(\mathbb{T}^{2n_0})$ and $\pi_* m_1 = \mu_1$, $\pi_* m = \mu_2$. We have by the properties of the pushforward

$$\begin{aligned} \lim_{t \rightarrow 0} \frac{\tilde{U}((1-t)m_1 + tm_2) - \tilde{U}(m_1)}{t} &= \lim_{t \rightarrow 0} \frac{U((1-t)\mu_1 + t\mu_2) - U(\mu_2)}{t} \\ &= \int_{\mathbb{T}^{n_0}} \frac{\delta U}{\delta \mu}(\mu_1, x_1)(\mu_2 - \mu_1)(dx_1) \\ &= \int_{\mathbb{T}^{n_0}} \frac{\delta U}{\delta \mu}(\pi_* m_1, x_1)(\pi_* m_2 - \pi_* m_1)(dx_1, dx_2) \\ &= \int_{\mathbb{T}^{n_0}} \frac{\delta U}{\delta \mu}(\pi_* m_1, x_1)(\pi_* m_2 - \pi_* m_1)(dx_1) \\ &= \int \int_{\mathbb{T}^{n_0} \times \mathbb{T}^{n_0}} \frac{\delta U}{\delta \mu}(\pi_* m_1, \pi(x_1, x_2))(m_2 - m_1)(dx_1, dx_2). \end{aligned}$$

Equation (2.2.1) follows. The regularity of $\tilde{U}$ is a direct implication of (3.5.1) and the fact that U is C^1 . $\square$

3.6 Technical proofs of the chapter

In this section we provide the proofs of some Propositions which were too long and would decrease the readability of the paper. We recall assumption **(A')** and we consider the optimal control problem

$$U(t_0, m_0) := \inf_{a \in \mathcal{A}} \left\{ \int_{t_0}^T \int_{\mathbb{T}^n} L(x, \alpha(x, t)) m_t(dx) dt + \int_{t_0}^T \mathcal{F}(m(s)) ds + \mathcal{G}(m(T)) \right\}, \quad (3.6.1)$$

where $(m_t)_{t \in [t_0, T]} \in \mathcal{P}(\mathbb{T}^n)$ satisfies (in the weak sense) the Fokker-Planck equation

$$\begin{cases} -\partial_t m + \operatorname{tr}(\sigma \sigma^\top D^2 m) - \operatorname{div}(m b(x, \alpha)) = 0, \\ m(t_0) = m_0 \in \mathcal{P}(\mathbb{T}^d), \end{cases} \quad (3.6.2)$$

and where the infimum is over all $\alpha \in \mathcal{A}$, where $\mathcal{A}$ is the set of all admissible controls $\{\alpha : [0, T] \times \mathbb{T}^n \rightarrow A \mid \alpha \in L^2([0, T] \times \mathbb{T}^n, dt \otimes m_t; \mathbb{R}^n)\}$.

Proposition 3.6.1. *Under assumptions (A') from section 3.3, U , as defined in (3.6.1), is a viscosity solution of (3.3.1) in the sense of Definition 3.3.5.*

Proof. The proof is known in the literature and can be found in [SY24] (or in [CGK⁺23, Theorem 5.5] when we work in $\mathbb{R}^n$ instead of $\mathbb{T}^n$) among others, but we include it for completeness. We first note that due to [CGK⁺24, Proposition 2.5], the function U defined in (3.6.1) is continuous.

For the subsolution property, let $\varphi \in C^{1,2}$ such that $U - \varphi$ achieves its maximum at (t_0, m_0) with $t_0 < T$; the case $t_0 = T$ being trivial. By the dynamic programming principle we have for any control α and $h \in (0, T - t_0]$

$$U(t_0, m_0) \leq \int_{t_0}^{t_0+h} \int_{\mathbb{T}^n} L(x, \alpha(s, x)) m_s(dx) ds + \int_{t_0}^{t_0+h} \mathcal{F}(m(s)) ds + U(t_0 + h, m(t_0 + h)),$$

where $(m_s)_{s \in [t_0, t_0+h]}$ solves the Fokker-Planck equation driven by α with initial condition $m(t_0) = m_0$. Since (t_0, m_0) is a position of maximum, the above implies

$$\varphi(t_0, m_0) - \varphi(t_0 + h, m(t_0 + h)) \leq \int_{t_0}^{t_0+h} \int_{\mathbb{T}^n} L(x, \alpha(s, x)) m_s(dx) ds + \int_{t_0}^{t_0+h} \mathcal{F}(m(s)) ds.$$

We divide by h and let $h \rightarrow 0$. Then, by Itô's formula (see [CD⁺18b]) we deduce

$$\begin{aligned}
-\partial_t \varphi(t_0, m_0) - \int_{\mathbb{T}^n} \text{tr} \left(\sigma \sigma^\top D_x D_m \varphi(t_0, m_0, x) \right) m_0(dx) \\
- \int_{\mathbb{T}^n} b(x, \alpha(t_0, x)) \cdot D_m \varphi(t_0, m_0, x) m_0(dx) \\
\leq \int_{\mathbb{T}^n} L(x, \alpha(t_0, x)) m_0(dx) + \mathcal{F}(m_0).
\end{aligned}$$

The subsolution property follows by taking supremum over $\alpha \in \mathcal{A}$ in the above inequality.

To prove that U is a supersolution of (3.3.1), let $\varphi \in C^{1,2}$ and suppose that $\min \{U - \varphi\}$ is achieved at (t_0, m_0) . Since $U(T, m_0) = \mathcal{G}(m_0)$, we only need to examine the case where $t_0 < T$. Without loss of generality, we may also assume that $U \geq \varphi$ and that the maximum at (t_0, m_0) is strict. We argue by contradiction and we suppose that

$$\begin{aligned}
-\partial_t \varphi(t_0, m_0) - \int_{\mathbb{T}^n} \text{tr} \left(\sigma \sigma^\top D_x D_m \varphi(t_0, m_0, x) \right) m_0(dx) \\
+ \int_{\mathbb{T}^n} H(x, D_m \varphi(t_0, m_0, x)) m_0(dx) < \mathcal{F}(m_0).
\end{aligned}$$

The regularity of φ allows us to choose an $\varepsilon > 0$ and an open neighbourhood B of (t_0, m_0) in $[t_0, T] \times \mathcal{P}(\mathbb{T}^n)$ such that

$$\begin{aligned}
\varepsilon - \partial_t \varphi(t, m) - \int_{\mathbb{T}^n} \text{tr} \left(\sigma \sigma^\top D_x D_m \varphi(t, m, x) \right) m(dx) \\
+ \int_{\mathbb{T}^n} H(x, D_m \varphi(t, m, x)) m(dx) < \mathcal{F}(m),
\end{aligned}$$

for every $(t, m) \in B$. For any admissible control $\alpha \in \mathcal{A}$, let $(m_s^\alpha)_{s \in [t_0, T]}$ be the solution to the Fokker-Planck equation (3.6.2) driven by α with initial condition $m(t_0) = m_0$. We also consider

$$t^\alpha := \inf_{\alpha \in \mathcal{A}} \{s \in [t_0, T] \mid (s, m_s^\alpha) \notin B\}. \quad (3.6.3)$$

By Itô's formula (see [CD⁺18b]) we have by our hypothesis

$$\begin{aligned}
\varphi(t^\alpha, m_{t^\alpha}^\alpha) &= \varphi(t_0, m_0) + \int_{t_0}^{t^\alpha} \partial_t \varphi(s, m_s^\alpha) ds \\
&\quad + \int_{t_0}^{t^\alpha} \int_{\mathbb{T}^n} b(x, \alpha(s, x)) \cdot D_m \varphi(s, m_s^\alpha, x) m_s(dx) ds \\
&\quad + \int_{t_0}^{t^\alpha} \int_{\mathbb{T}^n} \text{tr} \left(\sigma \sigma^\top D_x D_m \varphi(s, m_s^\alpha, x) \right) m_s(dx) ds \\
&> \varphi(t_0, m_0) - \int_{t_0}^{t^\alpha} \mathcal{F}(m(s)) ds - \int_{t_0}^{t^\alpha} \int_{\mathbb{T}^n} L(x, \alpha(s, x)) m_s(dx) ds + \varepsilon(t^\alpha - t_0)
\end{aligned}$$

or

$$\int_{t_0}^{t^\alpha} \left(\mathcal{F}(m(s)) + \int_{\mathbb{T}^n} L(x, \alpha(s, x)) m_s(dx) \right) ds + U(t^\alpha, m_{t^\alpha}^\alpha) > U(t_0, m_0) + h(\alpha),$$

where $h(\alpha) = U(t^\alpha, m_{t^\alpha}^\alpha) - \varphi(t^\alpha, m_{t^\alpha}^\alpha) + \varepsilon(t^\alpha - t_0)$. We take the infimum over α in the above inequality and by the dynamic programming principle we deduce $0 \geq \inf_\alpha h(\alpha)$. However, since $(t^\alpha, m_{t^\alpha}^\alpha) \in \partial B$, we have

$$h(\alpha) > \inf_{\partial B} \{U(t, m) - \varphi(t, m)\} + \varepsilon(t^\alpha - t_0) \Rightarrow \inf_\alpha h(\alpha) \geq \inf_{\partial B} \{U(t, m) - \varphi(t, m)\} > 0,$$

by the strict maximum property of (t_0, m_0) ; contradiction. $\square$

For the proof of the comparison principle, we are going to need the following Lemma.

Lemma 3.6.2. *Assume (A') from section 3.3 and let $n \in \mathbb{N}$ and $\alpha : [0, T] \times \mathbb{T}^n \rightarrow \mathbb{R}^n$ be a smooth control. Then, the function $V : [0, T] \times \mathcal{P}(\mathbb{T}^n) \rightarrow \mathbb{R}$ with*

$$V(t, m) = \int_t^T \int_{\mathbb{T}^n} L(x, \alpha(s, x)) m_s(dx) ds + \int_t^T \mathcal{F}(m(s)) ds + \mathcal{G}(m(T)),$$

where $(m_s)_{s \in [t, T]}$ solves the Fokker-Planck equation

$$\begin{cases} -\partial_s m + \text{tr}(\sigma \sigma^\top D^2 m) + \text{div}(mb(x, \alpha(s, x))) = 0, \\ m(t) = m, \end{cases}$$

in the weak sense, is a smooth supersolution of (3.3.1).

Proof. Since $m \in \mathcal{P}(\mathbb{T}^n)$, by Remark 3.5.2 (iii), $m \in \mathcal{P}_2(\mathbb{T}^n)$. Let $\xi \in L^2(\mathbb{T}^n)$ with $\mathcal{L}(\xi) = m$. By the smoothness of α and the Lipschitz continuity of b , we know that the Fokker-Planck equation expresses the evolution law of the process satisfying the SDE

$$X_r^m = \xi + \int_t^r b(X_s^m, \alpha(s, X_s^m)) ds + \sqrt{2}\sigma(B_r - B_t), \quad (3.6.4)$$

where B is Brownian motion. Then, we have

$$V(t, m) = \mathbb{E} \left[\int_t^T L(X_s^m, \alpha(s, X_s^m)) ds \right] + \int_t^T \mathcal{F}(\mathcal{L}(X_s^m)) ds + \mathcal{G}(\mathcal{L}(X_T^m))$$

and the smoothness of V follows by the regularity of L , $\mathcal{F}$ and $\mathcal{G}$ and the results of [BDL17]. For the supersolution property we have for $h \in (0, T - t]$ by the definition of V and Itô's formula (we refer to [CD⁺18b])

$$\begin{aligned} \frac{V(t+h, m) - V(t, m)}{h} &= \frac{V(t+h, m(t+h)) - V(t, m)}{h} - \frac{V(t+h, m(t+h)) - V(t+h, m)}{h} \\ &= -\frac{1}{h} \int_t^{t+h} \left(\int_{\mathbb{T}^n} L(x, \alpha(s, x)) m_s(dx) + \mathcal{F}(m(s)) \right) ds \\ &\quad - \frac{1}{h} \int_t^{t+h} \int_{\mathbb{T}^n} b(x, \alpha(s, x)) \cdot D_m V(s, m_s, x) m_s(dx) ds \\ &\quad - \frac{1}{h} \int_t^{t+h} \int_{\mathbb{T}^n} \text{tr}(\sigma \sigma^\top D_x D_m V(s, m_s, x)) m_s(dx) ds. \end{aligned}$$

We send $h \rightarrow 0$ and we discover

$$\begin{aligned} \partial_t V(t, m) = & - \int_{\mathbb{T}^n} L(x, \alpha(t, x)) m(dx) - \mathcal{F}(m) - \int_{\mathbb{T}^n} b(x, \alpha(t, x)) D_m V(t, m, x) m(dx) \\ & - \int_{\mathbb{T}^n} \text{tr} \left(\sigma \sigma^\top D_x D_m V(t, m, x) \right) m(dx). \end{aligned}$$

The supersolution property now follows from the inequality

$$L(x, \alpha(t, x)) + b(x, \alpha(t, x)) \cdot D_m V(t, m, x) \geq -H(x, D_m V(t, m, x)).$$

□

Proposition 3.6.3. *Assume (A3) with $n = 2d$ and (A'). Let U be the function defined in (3.6.1), U_1 be a viscosity subsolution of (3.3.1) and U_2 be a viscosity supersolution of (3.3.1), then for every $(t, m) \in [0, T] \times \mathcal{P}(\mathbb{T}^n)$ we have $U_1(t, m) \leq U(t, m) \leq U_2(t, m)$.*

Proof. We start by showing that $U_2 \geq U$. To this end, for $N \in \mathbb{N}$ and $\mathbf{x}_0 \in (\mathbb{T}^n)^N$, we consider the N -particle system $\mathbf{X}_t = (X_t^1, \dots, X_t^N)$ such that

$$X_t^k = x_0^k + \int_0^t b(X_s^k, \alpha_s^k) ds + \sqrt{2} \sigma dB_s^k, \quad k = 1, \dots, N, \quad (3.6.5)$$

where $\alpha = (\alpha^1, \dots, \alpha^N)$ is an admissible control and B^k , $k = 1, \dots, N$ are independent Brownian motions. For $t_0 \in [0, T]$, we also write the following control problem associated with (3.6.5)

$$\mathcal{V}^N(t_0, \mathbf{x}_0) := \inf_{\alpha \in (\mathcal{A})^N} \mathbb{E} \left[\int_{t_0}^T \left(\frac{1}{N} \sum_{k=1}^N L(X_t^k, \alpha_t^k) + \mathcal{F}(m_{\mathbf{X}_t}^N) \right) dt + \mathcal{G}(m_{\mathbf{X}_T}^N) \right], \quad (3.6.6)$$

where $m_{\mathbf{x}}^N = \frac{1}{N} \sum_{k=1}^N \delta_{x^k}$ is the empirical measure. Furthermore, for $m \in \mathcal{P}_2(\mathbb{T}^n)$, we set

$$\begin{aligned}\widehat{\mathcal{V}}^N(t, m) &= \int_{(\mathbb{T}^n)^N} \mathcal{V}^N(t, \mathbf{x}) \prod_{k=1}^N m(dx^k), \quad \widehat{\mathcal{F}}^N(m) = \int_{(\mathbb{T}^n)^N} \mathcal{F}(m_{\mathbf{x}}^N) \prod_{k=1}^N m(dx^k) \\ \text{and } \widehat{\mathcal{G}}^N(m) &= \int_{(\mathbb{T}^n)^N} \mathcal{G}(m_{\mathbf{x}}^N) \prod_{k=1}^N m(dx^k).\end{aligned}$$

By [CDJS23, Lemma 3.6] and [CM20, Proposition 3.1]⁷, we know that $\widehat{\mathcal{V}}^N \in C^{1,2}$ and satisfies

$$\begin{cases} -\partial_t \widehat{\mathcal{V}}^N(t, m) - \int_{\mathbb{T}^n} \text{tr}(\sigma \sigma^\top D_m \widehat{\mathcal{V}}^N(t, m, y)) m(dy) \\ \quad + \int_{\mathbb{T}^n} H(x, D_m \widehat{\mathcal{V}}^N(t, m, x)) m(dx) \leq \widehat{\mathcal{F}}^N(m), & \text{in } [0, T) \times \mathcal{P}_2(\mathbb{T}^n), \\ \widehat{\mathcal{V}}^N(T, m) = \widehat{\mathcal{G}}^N(m), & \text{in } \mathcal{P}_2(\mathbb{T}^n). \end{cases} \quad (3.6.7)$$

Let $M > 0$ and $A_M = \{m \in \mathcal{P}(\mathbb{T}^{n_0}) | M_2(m) \leq M\}$. It suffices to show that $U \leq U_2$ in $Q_T^M := [0, T] \times A_M$. If the opposite is true, then, there exist ε, δ such that $\max_{Q_T^M} \{U - U_2 - \varepsilon(T - t)\} > 2\delta$. However, by the proof of [CDJS23, Proposition 3.7], we have

$$\left| \widehat{\mathcal{V}}^N(t, m) - U(t, m) \right| \leq \frac{C}{N^\beta} (1 + M_2^{1/2}(m)) \leq \frac{C}{N^\beta} (1 + M^{1/2}),$$

for every $(t, m) \in Q_T^M$, for some constants $C, \beta > 0$, therefore by choosing N large enough

$$\max_{(t, m) \in Q_T^M} \left\{ \widehat{\mathcal{V}}^N(t, m) - U_2(t, m) - \varepsilon(T - t) \right\} > \delta. \quad (3.6.8)$$

Since $\widehat{\mathcal{V}}^N$ is continuous and U_2 is lower semi-continuous, the maximum in (3.6.8) is attained at a point $(t_0, m_0) \in Q_T^M$. We distinguish two cases: a) $t_0 = T$ and b) $t_0 < T$.

If $t_0 = T$, then $\widehat{\mathcal{G}}^N(m_0) > \delta + U_2(T, m_0)$, which by the supersolution property of U_2 implies

⁷ In these reference $\widehat{\mathcal{V}}^N$ is defined on $\mathcal{P}(\mathbb{R}^n)$ instead of the torus, but the proof still works in the case of $\mathcal{P}(\mathbb{T}^n)$.

$\widehat{\mathcal{G}}^N(m_0) > \delta + \mathcal{G}(m_0)$. But, since $\mathcal{G}$ is Lipschitz continuous it follows from [FG15, Theorem 1] that

$$\left| \widehat{\mathcal{G}}^N(m_0) - \mathcal{G}(m_0) \right| \leq C \int_{(\mathbb{T}^n)^N} \mathbf{d}_1(m_{\mathbf{x}}^N, m_0) \prod_{i=1}^N m_0(dx^i) \leq \frac{C}{N^\beta} (1 + M^{1/2}),$$

thus, by choosing N large enough we get a contradiction with $\widehat{\mathcal{G}}^N(m_0) > \delta + \mathcal{G}(m_0)$.

If $t_0 < T$, then (t_0, m_0) is a point of minimum for $U_2 - (\widehat{\mathcal{V}}^N - \varepsilon(T - t))$, so combining the definition of viscosity supersolution with (3.6.7), we get $|\widehat{\mathcal{F}}^N(m_0) - \mathcal{F}(m_0)| \geq \varepsilon$. We can arrive at a contradiction if we choose N large enough by a similar argument as in the previous case (with $\mathcal{F}$ instead of $\mathcal{G}$).

Now we show that $U_1 \leq U$. Let's assume that $\max\{U_1 - U\} > 0$ and, by the upper semi-continuity of $U_1 - U$, that the maximum is achieved at a point (t_0, m_0) . For $\varepsilon > 0$, we consider $\alpha^\varepsilon \in \mathcal{A}$ to be an ε -optimal smooth control for $U(t_0, m_0)$ in (3.6.1) and for $(t, m) \in [0, T] \times \mathcal{P}(\mathbb{T}^n)$ we introduce

$$U^\varepsilon(t, m) = \int_t^T \int_{\mathbb{T}^n} L(x, \alpha^\varepsilon(s, x)) m_s(dx) ds + \int_t^T \mathcal{F}(m(s)) ds + \mathcal{G}(m(T)), \quad (3.6.9)$$

where $(m_s)_{s \in [t, T]}$ satisfies (3.6.2) with $\alpha = \alpha^\varepsilon$, $t_0 = t$ and initial condition m (at t). By the ε -optimality of α^ε , we know that $U^\varepsilon(t_0, m_0) \xrightarrow{\varepsilon \rightarrow 0} U(t_0, m_0)$, so we may choose $\varepsilon > 0$ small enough such that

$$\max_{t, m} \left\{ U_1(t, m) - U^\varepsilon(t, m) - \varepsilon(T - t) \right\} > 0. \quad (3.6.10)$$

Let $(t^\varepsilon, m^\varepsilon)$ be the point where the maximum in (3.6.10) is achieved. We distinguish two cases: a) $t^\varepsilon = T$ and b) $t^\varepsilon < T$.

If $t^\varepsilon = T$, then $U_1(T, m^\varepsilon) > U^\varepsilon(T, m^\varepsilon)$, hence $U_1(T, m^\varepsilon) > \mathcal{G}(m^\varepsilon)$, which contradicts the definition of a subsolution.

If $t^\varepsilon < T$, then, again by the definition of a subsolution for U^ε as the test function, we get

$$\begin{aligned} \varepsilon - \partial_t U^\varepsilon(t^\varepsilon, m^\varepsilon) + \int_{\mathbb{T}^n} H(x, D_m U^\varepsilon(t^\varepsilon, m^\varepsilon, x)) m^\varepsilon(dx) \\ - \int_{\mathbb{T}^n} \text{tr}(\sigma \sigma^\top D_x D_m U^\varepsilon(t^\varepsilon, m^\varepsilon, x)) m^\varepsilon(dx) \leq \mathcal{F}(m^\varepsilon). \end{aligned}$$

However, by Lemma 3.6.2, U^ε is a smooth supersolution of (3.3.1), which contradicts the above inequality. $\square$

Lemma 3.6.4. *Let $\delta > 0$, σ_2 an invertible $d \times d$ matrix and $f : \mathbb{R}^d \times \mathbb{T}^d \rightarrow \mathbb{R}^d$, $g : \mathbb{R}^d \times \mathbb{T}^d \rightarrow \mathbb{R}$ two C^1 functions. Suppose that $u := u(x, y) : \mathbb{R}^d \times \mathbb{T}^d$ satisfies the elliptic equation*

$$\delta u - \text{tr}(\sigma_2 \sigma_2^\top D_{yy}^2 u) + f(x, y) \cdot D_y u = g(x, y), \quad (3.6.11)$$

in the classical sense. Then, u is C^1 ; that is $D_x u(x, y)$ and $D_y u(x, y)$ exist and are continuous.

Proof. Fix $x \in \mathbb{R}^d$ and $B \subset \mathbb{R}^d$ a closed ball centered at x . In the proof we will be working in $B \times \mathbb{T}^d$, which is compact. Let $h \in \mathbb{R}^d$ such that $x + h \in B$, $\xi = \frac{h}{|h|}$ and $\tilde{u}_h(x, y) = u(x + h, y) - u(x, y)$. The function $\tilde{u}_h$ satisfies

$$\begin{aligned} \delta \tilde{u}_h - \text{tr}(\sigma_2 \sigma_2^\top D_{yy}^2 \tilde{u}_h) + f(x + h, y) \cdot D_y \tilde{u}_h \\ = g(x + h, y) - g(x, y) - (f(x + h, y) - f(x, y)) \cdot D_y u(x, y). \end{aligned} \quad (3.6.12)$$

By the compactness of $B \times \mathbb{T}^d$, we have, for some constant $C > 0$, $\|D_y u(x, \cdot)\|_{L^\infty} \leq C$ and $|f(x + h, y)| \leq \sup_{(z, y) \in B \times \mathbb{T}^d} |f(z, y)|$. In addition, for any $y \in \mathbb{T}^d$

$$|g(x + h, y) - g(x, y)| \leq |h| \int_0^1 |D_x g(sh + x, y)| ds \leq |h| \sup_{(z, y) \in B \times \mathbb{T}^d} |D_x g(z, y)|,$$

$$|f(x + h, y) - f(x, y)| \leq |h| \sup_{(z, y) \in B \times \mathbb{T}^d} |D_x f(z, y)|.$$

By virtue of these bounds and the maximum principle, we have that there exists a constant $C > 0$ depending only on $\sigma_2, \delta, d, f, g$ and B such that $\|\tilde{u}_h(x, \cdot)\|_{L^\infty} \leq C$. Now, standard Hölder estimates for elliptic equations (see [GT77, Chapter 8]), yield

$$\|\tilde{u}_h\|_{C^{1,\alpha}} \leq C(\|\tilde{u}_h\|_{L^\infty} + C'|h|) \leq C|h|, \quad (3.6.13)$$

for some $\alpha \in (0, 1)$ and where C depends only on $\delta, d, \sigma_2, f, g$ and B . On the one hand (3.6.13) implies that $D_y u(x, y)$ is continuous. On the other hand, it allows us to use Arzela-Ascoli's theorem to get a uniformly convergent subsequence of $\tilde{u}_h(x, \cdot)/|h|$ in C^1 as $|h| \rightarrow 0$. Let $\bar{u}_\xi$ be the limit, which is continuous due to the uniform convergence. Since $\tilde{u}_h$ is a classical solution of (3.6.12), it is also a weak solution, hence $\bar{u}_\xi$ is a weak solution of

$$\delta \bar{u}_\xi - \text{tr} \left(\sigma_2 \sigma_2^\top D_{yy}^2 \bar{u}_\xi \right) + f(x, y) \cdot D_y \bar{u}_\xi = D_x g(x, y) \cdot \xi - D_x f(x, y) \xi \cdot D_y u(x, y). \quad (3.6.14)$$

Since the weak solution is unique, we deduce that $\frac{\tilde{u}_h(x, \cdot)}{|h|} \xrightarrow{|h| \rightarrow 0} \bar{u}_\xi$, therefore the directional derivatives of $u(x)$ exist and hence $D_x u(x, y)$ exists. From (3.6.13) by sending $|h| \rightarrow 0$, it also follows that $D_x u(x, \cdot) \in C^{1,\alpha}(\mathbb{T}^d)$ with a Hölder constant which is uniform in B (since $x \in B$).

To prove the continuity of $D_x u(x, y)$, since $D_x u(x, \cdot) \in C^{1,\alpha}(\mathbb{T}^d)$ with a Hölder constant which is uniform in B , it suffices to show that $x \mapsto D_x u(x, y)$ is continuous in B . We will show that its components $u_{x_i}(x, y)$, $i = 1, \dots, d$, are continuous with respect to x . Indeed, let $x \in B$ and $h \in \mathbb{R}^d$ such that $x + h \in B$. We consider the difference $\bar{w}_i^h(x, y) := u_{x_i}(x + h, y) - u_{x_i}(x, y)$. By choosing the appropriate $\xi = (0, \dots, 0, 1, 0, \dots, 0)$, where 1 is in the i -th position in (3.6.14), we discover that $\bar{w}_i^h$ satisfies in the weak sense

$$\begin{aligned} \delta \bar{w}_i^h(y) - \text{tr} \left(\sigma_2 \sigma_2^\top D_{yy}^2 \bar{w}_i^h(y) \right) + f(x + h, y) \cdot D_y \bar{w}_i^h(y) \\ = g_{x_i}(x + h, y) - g_{x_i}(x, y) - f_{x_i}(x + h, y) \cdot D_y u(x + h, y) + f_{x_i}(x, y) \cdot D_y u(x, y) \end{aligned}$$

$$- (f(x+h, y) - f(x, y)) \cdot D_y u_{x_i}(x, y). \quad (3.6.15)$$

Since $D_x u(x, \cdot) \in C^{1,\alpha}(\mathbb{T}^d)$ and $f, g \in C^1$, we observe that the coefficients on the left hand side and the right hand side of (3.6.15) are uniformly bounded (with respect to h) as functions of y , hence again by standard Hölder estimates $\bar{w}_i^h$ is uniformly bounded in $C^{1,\alpha}$. By Arzela-Ascoli, $\bar{w}_i^h$ converges uniformly (up to a subsequence) as $|h| \rightarrow 0$ to a function $\bar{w}_i$. Since the right hand side of (3.6.15) converges to 0 uniformly as $|h| \rightarrow 0$, $\bar{w}$ satisfies weakly the equation

$$\delta \bar{w}_i - \text{tr} \left(\sigma_2 \sigma_2^\top D_{yy}^2 \bar{w}_i \right) + f(x, y) \cdot D_y \bar{w}_i = 0.$$

We notice that $\bar{w}_i = 0$ is the unique weak solution to this equation, hence $\bar{w}_i^h \xrightarrow{|h| \rightarrow 0} 0$ uniformly. We conclude that $\lim_{|h| \rightarrow 0} \|u_{x_i}(x+h, \cdot) - u_{x_i}(x, \cdot)\|_{L^\infty} = 0$, which gives us the desired continuity.

□

We, finally, prove a regularization result for functions defined on the Wasserstein space $\mathcal{P}_p(\mathbb{T}^d)$, $p \geq 1$. The idea comes from [CD⁺18b, Chapter 5].

Lemma 3.6.5. *Let $\Phi : \mathcal{P}_p(\mathbb{T}^d) \rightarrow \mathbb{R}$ be a function, which is Lipschitz and continuously differentiable. For a smooth symmetric density on $\mathbb{R}^d$ with compact support and $\delta > 0$, let*

$$\rho_\delta(x) := \frac{1}{\delta^d} \rho\left(\frac{x}{\delta}\right), \quad x \in \mathbb{R}^d$$

and $\Phi^\delta(m) := \Phi(m * \rho_\delta)$, $m \in \mathcal{P}_p(\mathbb{T}^d)$. Then, Φ^δ is continuously differentiable with

$$\frac{\delta \Phi^\delta}{\delta m}(m, x) = \left(\frac{\delta \Phi}{\delta m}(m * \rho_\delta, \cdot) * \rho_\delta \right)(x), \quad x \in \mathbb{T}^d, \quad m \in \mathcal{P}_p(\mathbb{T}^d).$$

Moreover, if $\frac{\delta \Phi}{\delta m}(m, \cdot) \in C^2$, then there exist a positive constant $C := C(d, p, \rho)$ and an

increasing function $\omega : [0, +\infty) \rightarrow [0, +\infty)$ with $\lim_{t \rightarrow 0^+} \omega(t) = \omega(0) = 0$ such that

$$\sup_{m \in \mathcal{P}_p(\mathbb{T}^d)} |\Phi^\delta(m) - \Phi(m)| + \sup_{m \in \mathcal{P}_p(\mathbb{T}^d)} \left\| \frac{\delta \Phi^\delta}{\delta m}(m, \cdot) - \frac{\delta \Phi}{\delta m}(m, \cdot) \right\|_{C^2(\mathbb{T}^d)} \leq C\omega(\delta) \quad (3.6.16)$$

Proof. The proof for the derivative can be found in [CD⁺18b, Chapter 5] and the proof for the first term of (3.6.16) follows from the Lipschitz continuity of Φ as in [DDJ24, Lemma 4.2]. Indeed, for $m \in \mathcal{P}_p(\mathbb{T}^d)$ and L_Φ the Lipschitz constant of Φ

$$\begin{aligned} |\Phi^\delta(m) - \Phi(m)| &\leq L_\Phi \mathbf{d}_p(m * \rho_\delta, m) \leq L_\Phi \mathbf{d}_1(m * \rho_\delta, m)^{1/p} \\ &\leq L_\Phi \left(\sup_{f \in 1\text{-Lip}} \int_{\mathbb{T}^d} (f * \rho_\delta(x) - f(x)) m(dx) \right)^{1/p} \leq L_\Phi \delta^{1/p}. \end{aligned}$$

For the second term of (3.6.16) we argue similarly. Since the functions $D_x D_m \Phi(x, m)$ and $D_m \Phi(x, m)$ are continuous, by the compactness of $\mathbb{T}^d \times \mathcal{P}_p(\mathbb{T}^d)$, they must also be uniformly continuous, hence we may consider $\omega_{D_m \Phi}$ and $\omega_{D_x D_m \Phi}$ the corresponding moduli of continuity. We have for any $(m, x) \in \mathcal{P}_p(\mathbb{T}^d) \times \mathbb{T}^d$

$$\begin{aligned} |D_m \Phi(m * \rho_\delta, \cdot) * \rho_\delta(x) - D_m \Phi(m, x)| &\leq |D_m \Phi(m * \rho_\delta, \cdot) * \rho_\delta(x) - D_m \Phi(m, \cdot) * \rho_\delta(x)| \\ &\quad + |D_m \Phi(m, \cdot) * \rho_\delta(x) - D_m \Phi(m, x)|. \end{aligned}$$

However,

$$\begin{aligned} |D_m \Phi(m, \cdot) * \rho_\delta(x) - D_m \Phi(m, x)| &\leq \int_{\mathbb{T}^d} |D_m \Phi(m, x - y) - D_m \Phi(m, x)| \rho_\delta(y) dy \\ &\leq \int_{\mathbb{T}^d} \omega_{D_m \Phi}(|y|) \rho_\delta(y) dy \leq C\omega_{D_m \Phi}(\delta). \end{aligned}$$

and

$$|D_m \Phi(m * \rho_\delta, \cdot) * \rho_\delta(x) - D_m \Phi(m, \cdot) * \rho_\delta(x)|$$

$$\begin{aligned}
&\leq \int_{\mathbb{T}^d} |D_m \Phi(m * \rho_\delta, y) - D_m \Phi(m, y)| \rho_\delta(x - y) dy \\
&\leq \int_{\mathbb{T}^d} \omega_{D_m \Phi}(\mathbf{d}_p(m * \rho_\delta, m)) \rho_\delta(x - y) dy \\
&\leq \int_{\mathbb{T}^d} \omega_{D_m \Phi}(\delta^{1/p}) \rho_\delta(x - y) dy \\
&\leq C \omega_{D_m \Phi}(\delta^{1/p}).
\end{aligned}$$

Thus,

$$|D_m \Phi(m * \rho_\delta, \cdot) * \rho_\delta(x) - D_m \Phi(m, x)| \leq C \max \left\{ \omega_{D_m \Phi}(\delta^{1/p}), \omega_{D_m \Phi}(\delta) \right\}.$$

This shows the inequality for the first derivative. For the second derivative, we argue similarly but with the modulus $\omega_{D_x D_m \Phi}$. The proof is complete. $\square$

Acknowledgements. The author was partially supported by P. E. Souganidis' NSF grant DMS-1900599, ONR grant N000141712095, and AFOSR grant FA9550-18-1-0494. The author wishes to thank Joe Jackson for insightful discussions regarding Theorem 3.3.6 and its proof.

CHAPTER 4

CALIBRATION OF STOCHASTIC VOLATILITY MODELS VIA MARTINGALE OPTIMAL TRANSPORT AND MARTINGALE SCHRÖDINGER BRIDGES

4.1 Introduction

Since the introduction of the Black–Scholes model [BS73], a significant amount of effort has been devoted to developing sophisticated volatility models that properly capture the market dynamics. To correct for any systematic errors and ensure accurate predictions, it is necessary to properly adjust these models to Vanilla options¹, such as an S&P500 (SPX) option, which depends on the eponymous stock market index², and VIX options (options on the volatility index). These problems are called calibration problems and are the subject of intense current research; we refer to the survey [GLOW21] and the references therein. To calibrate stochastic volatility models, many authors have employed the tools of optimal transport and Schrödinger bridges, which are related to the transportation of a probability measure to another in an optimal way. The optimization is with respect to a given cost functional and the relative entropy, respectively. Moreover, due to the arbitrage-free nature of many studies, additional martingale constraints are considered. We give an overview of these problems and outline their role in the calibration of stochastic volatility models.

1. The term Vanilla options refers to financial contracts that give the holder the right to buy or sell an asset at a predetermined price within a specific time frame.

2. The Standard and Poor’s 500 index, or simply the S&P 500, tracks the performance of 500 of the largest publicly traded companies in the United States.

4.1.1 Overview

Let $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R})$ be two Borel probability measures in convex order³ $\mu_1 \leq_c \mu_2$ and $T_2 > T_1 > 0$. In the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}_0)$, with $\Omega := C([0, T_2]; \mathbb{R}^2)$, we consider the continuous time stochastic volatility model

$$\begin{cases} dX_t = \sigma(X_t, Y_t) dW_t, \\ dY_t = b(X_t, Y_t) dt + \tau_1(X_t, Y_t) dW_t + \tau_2(X_t, Y_t) dW_t^\perp, \end{cases} \quad (4.1.1)$$

where $W, W^\perp$ are independent $\mathbb{P}_0$ -Brownian motions and $\sigma, b, \tau_1, \tau_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$. In the mathematical finance literature, known stochastic volatility models captured by this general form can be the Heston model and the SABR model. In (4.1.1), X_t represent the price of S&P500, while Y_t is the value of its volatility. Calibrating (4.1.1) to SPX options, is equivalent to finding a probability measure $\mathbb{P}$ in $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t \geq 0})$, such that the laws $\mathcal{L}^\mathbb{P}(X_{T_1}) = \mu_1$ and $\mathcal{L}^\mathbb{P}(X_{T_2}) = \mu_2$ ⁴. At the same time, we require X_t to be a $\mathbb{P}$ -martingale and $\mathbb{P}$ to be close to $\mathbb{P}_0$, so that the models remain as close as possible to the initial belief (4.1.1).

In many works, the role of martingale optimal transport and martingale Schrödinger bridges is crucial, especially in quantifying how close $\mathbb{P}$ will be to $\mathbb{P}_0$. In particular, it is needed to establish the probability measure $\mathbb{P}$ which is closest to $\mathbb{P}_0$ in the appropriate sense (martingale optimal transport or martingale Schrödinger bridges). These minimization problems often admit duality formulas, which are important in the numerical approximation of the optimal $\mathbb{P}^*$, whenever it exists. We refer to [GLW17, GLOW22, Guy22] for numerical simulations using the duality formulas.

3. This means that $\int h(x) d\mu_1(x) \leq \int h(x) d\mu_2(x)$ for any convex function $h : \mathbb{R} \rightarrow \mathbb{R}$.

4. In this case T_1, T_2 are the maturities of the options, while the constraints on law give the prices of the options with maturities T_1, T_2 : $\mathbb{E}^\mathbb{P}[(X_{T_i} - a)_+] = \int (x - a)_+ d\mu_i(x)$, $i = 1, 2$, for all strike prices $a \in \mathbb{R}$.

4.1.2 Calibration via martingale optimal transport

Let $T_0 \in [0, T_2]$, $\mu_0 \in \mathcal{P}(\mathbb{R})$ be a probability measure and $c : \mathbb{R}^2 \rightarrow \mathbb{R}^+$ a given function. The problem of optimal transport was first addressed by Monge [Mon81] and later by Kantorovich [Kan48] and, in its form with μ, ν as marginals⁵, refers to

$$\inf \left\{ \mathbb{E}[c(X_{T_0}, X_{T_2})] : \mathcal{L}^{\mathbb{P}}(X_{T_0}) = \mu_0, \mathcal{L}^{\mathbb{P}}(X_{T_2}) = \mu_2 \right\} \quad (4.1.2)$$

where the minimization is over all random vectors (X_0, X_{T_2}) in a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, which are called transportation plans. For every transportation plan and $\omega \in \Omega$, the mass from position $X_0(\omega)$ is transported to position $X_{T_2}(\omega)$. In 2000, Benamou and Brenier [BB00] introduced a continuous in time formulation of this problem. The dual formulation of the time-continuous problem has been extensively studied as an application of the Fenchel-Rockafellar theorem [Vil21, Theorem 1.9].

Extending the Monge-Kantorovich problem and its Benamou-Brenier formulation, Mikami and Thieullen [MT06], and Tan and Touzi [TT13] introduced the continuous time semi-martingale optimal transport problem and its dual formulation. In the context of calibration problems with two maturities T_1, T_2 from the previous subsection, for $T_0 \in [0, T_2]$, $(X_t)_{t \in [T_0, T_2]}$ a continuous time stochastic process with values in $\mathbb{R}$ over the filtered space $(\Omega, \mathcal{F}, \mathbb{F})$, the semi-martingale optimal transport problem can be written as

$$V(T_0, \mu_0, \mu_1, \mu_2) = \inf_{\mathbb{P} \in \mathcal{P}_m} \left\{ \mathbb{E}^{\mathbb{P}} \left[\int_{T_0}^{T_2} L(\sigma_t \sigma_t^\top) ds \right] \right\}, \quad (4.1.3)$$

where $L : \mathbb{R} \rightarrow \mathbb{R}$ is a given function. The minimization is over the set of probability

5. See [Pas15] for a version of the Monge-Kantorovich problems with more than two marginals.

measures in $(\Omega, \mathcal{F}, \mathbb{F})$

$$\mathcal{P}_m = \left\{ \mathbb{P} \left| X_t \text{ is a } \mathbb{P}\text{-semimartingale and } \mathcal{L}^{\mathbb{P}}(X_{T_i}) = \mu_i, i = 0, 1, 2 \right. \right\} \quad (4.1.4)$$

and X_t , being a $\mathbb{P}$ -semimartingale, satisfies $dX_t = \beta_t dt + \sigma_t dB_t$ with $(B_t)_{t \geq 0}$ being a 1-dimensional Brownian motion over the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ and β_t, σ_t are two adapted processes with values in $\mathbb{R}$. Problem (4.1.3) does not depend on an initial measure $\mathbb{P}_0$, hence it corresponds to a calibration problem for X_t with no initial belief $\mathbb{P}_0$.

The continuous time martingale optimal transport problem is equivalent to (4.1.3) with the additional constraint that X_t is a martingale, that is $\beta_t = 0$. It originates from questions of worst case bounds for derivative prices in model-independent finance, see e.g [GHLT14]. Unlike the Monge-Kantorovich optimal transport problem and its Benamou-Brenier formulation, the connection between the continuous time martingale optimal transport problem and its discrete analogue is still not well understood (see [HT19]). However, (4.1.3) still possesses a duality formula, which is useful in approximating the optimal $\mathbb{P}$ (e.g [GLW17, GLOW22] among others).

4.1.3 Calibration via martingale Schrödinger bridges

The Schrödinger problem was first studied by Schrödinger [Sch31, Sch32] and it involves minimizing the relative entropy H . Using the ideas developed on that problem (we refer to the survey [Léo13]) Labordère [HL19] introduced the notion of martingale Schrödinger bridges.

We consider $t_0 \in [0, T]$ and (4.1.1) with initial condition $(X_{t_0}, Y_{t_0}) = (x, y) \in \mathbb{R}^2$, for some $x, y > 0$. The martingale Schrödinger bridge is the measure that is closest to $\mathbb{P}_0$ in the sense of relative entropy, while the SVM under $\mathbb{P}$ will satisfy some given conditions including the martingality of X_t . In the notation introduced at the beginning of the section, the martingale

Schrödinger bridge problem that calibrated (4.1.1) is the minimization problem

$$D_{\mathbb{P}_0} := \inf_{\mathbb{P} \in \mathcal{C}_{t_0}(\mu_1, \mu_2)} H(\mathbb{P}|\mathbb{P}_0), \quad (4.1.5)$$

where H is the relative entropy between $\mathbb{P}$ and $\mathbb{P}_0$ ⁶ and

$$\mathcal{C}_{t_0}(\mu_1, \mu_2) = \begin{cases} \left\{ \mathbb{P} \in \mathcal{P} \left| \mathcal{L}^{\mathbb{P}}(X_{T_i}) = \mu_i, i = 1, 2 \text{ and } X \text{ is } ((\mathcal{F}_t), \mathbb{P})\text{-martingale} \right. \right\}, & \text{if } t_0 \in [0, T_1), \\ \left\{ \mathbb{P} \in \mathcal{P} \left| \mathcal{L}^{\mathbb{P}}(X_{T_2}) = \mu_2, X \text{ is } ((\mathcal{F}_t), \mathbb{P})\text{-martingale} \right. \right\}, & \text{if } t_0 \in [T_1, T_2]. \end{cases} \quad (4.1.6)$$

Apart from calibrating SVM to SPX options [HL19], martingale Schrödinger bridges were also used to solve the longstanding *joint S&P 500 - VIX smile calibration problem* [Guy22, Guy20], yielding the so called dispersion constrained martingale Schrödinger bridges. This was done by considering the problem

$$D'_{\mathbb{P}_0} := \inf_{\mathbb{P} \in \mathcal{C}_{t_0}(\mu_1, \mu_2), \text{ VIX}} H(\mathbb{P}|\mathbb{P}_0), \quad (4.1.7)$$

where $\mathcal{C}_{t_0}(\mu_1, \mu_2)$ was defined in (4.1.6), (X_t, Y_t) satisfy (4.1.1) and VIX is a constraint on the law of

$$V_{\mathbb{P}} = \mathbb{E}^{\mathbb{P}} \left[-\log(X_{T_2}) + \log(X_{T_1}) | \mathcal{F}_{T_1} \right], \quad \mathcal{L}^{\mathbb{P}}(V_{\mathbb{P}}) = \mu_3, \quad (4.1.8)$$

for some given probability distribution $\mu_3 \in \mathcal{P}(\mathbb{R})$.

In the above references, if we assume that there are no further constraints, the numerical construction of the martingale Schrödinger bridge or the dispersion constrained martingale

6. The relative entropy between two measures μ, ν is defined as follows. $H(\mu|\nu) = \begin{cases} \int \log \frac{d\mu}{d\nu} d\mu, & \text{if } \mu \ll \nu, \\ +\infty, & \text{otherwise.} \end{cases}$

Schrödinger bridge, as in martingale optimal transport, relies on a dual formulation of the problem.

4.1.4 Main results

Our aim in this note is to give new proofs to duality formulas for martingale optimal transport or martingale Schrödinger bridges already existing in the literature or prove such duality formulas in case they were written down as formal statements.

For martingale optimal transport we have the following duality formula. It is worth noting that a duality formula was also proved in [GLW22, GLW17] if the constraints for X_{T_1}, X_{T_2} are not on their law, but on their averages. Additionally, in the case $T_0 > T_1$, i.e. the intermediate constraint at time T_1 becomes irrelevant, our duality formula can also be found in [HT19], as an application of the Fenchel-Rockafellar theorem. Our main idea of proof is to use the Von-Neumann minimax principle (Theorem 4.6.1), instead of the Fenchel-Rockafellar theorem, as well as ideas from mean field control theory.

Theorem 4.1.1. *Suppose that (A1), (A2) are true. Let $\mu_0, \mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R})$ be in convex increasing order ($\mu_0 \leq_c \mu_1 \leq_c \mu_2$), $0 < T_1 < T_2$ and $T_0 \in [0, T_2]$. Then, (4.1.3) is admissible. Furthermore, whenever V is finite, the following duality formula holds*

$$V(T_0, \mu_0, \mu_1, \mu_2) = \sup_{u_1, u_2 \in Lip} \left\{ \int u(T_0, x) d\mu_0(x) - \int u_2(x) d\mu_2(x) - \mathbf{1}_{[0, T_1]}(T_0) \int u_1(x) d\mu_1(x) \right\}, \quad (4.1.9)$$

where the supremum is taken over all bounded Lipschitz functions u_1, u_2 , u is a viscosity solution, in the sense of Definition 4.3.1, of the Hamilton-Jacobi equation

$$\begin{cases} -\partial_t u + H\left(\frac{D^2 u}{2}\right) = \delta_{T_1}(t) u_1, & (t, x) \in [T_0, T_2] \times \mathbb{R}^d, \\ u(T_2, x) = u_2(x), & x \in \mathbb{R}^d, \end{cases} \quad (4.1.10)$$

and $H(a) = \sup_{b \geq 0} \{-ab - L(b)\}$, for any $a \in \mathbb{R}$, is the Hamiltonian. Moreover, the same formula holds if the supremum in (4.1.9) is taken over u, u_1, u_2 such that u is a classical super-solution of (4.1.10).

As a consequence of Theorem 4.1.1 the following Corollary holds

Corollary 4.1.2. *Suppose that (A1) and (A2) hold and that H is uniformly elliptic; that is, there exists a constant $\lambda > 0$ such that $H'(a) \leq -\lambda$ for every $a \in \mathbb{R}$. Then, the supremum in (4.1.9) is achieved at $u_1^*, u_2^* \in C_b^2$ if and only if the infimum in (4.1.3) is achieved when the diffusion of X_s is $\sigma_s = \left(-H' \left(\frac{D^2 u(s, X_s)}{2}\right)\right)^{1/2}$, where u satisfies (4.1.10) in the classical sense of Definition 4.3.1 for the functions $u_1^*, u_2^* \in C_b^2$.*

For calibration via martingale Schrödinger bridges our first aim is to give a new proof of the duality formula established in [HL19, Theorem 3.5], again by employing the Von-Neumann minimax principle. Before stating the result, we consider the generator of (4.1.1)

$$\mathcal{L}_{x,y}^0 = b(x, y) \partial_y + \frac{\sigma^2(x, y)}{2} \partial_{xx} + \frac{\tau_1^2(x, y) + \tau_2^2(x, y)}{2} \partial_{yy} + \sigma(x, y) \tau_1(x, y) \partial_{xy}. \quad (4.1.11)$$

Theorem 4.1.3. *Let $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R})$. We assume (A3) and that (4.1.5) is admissible. Then, for $t_0 \in [0, T_2]$ and $(X_t, Y_t)_{t \in [t_0, T_2]}$ satisfying (4.1.1) we have*

$$\begin{aligned} & \inf_{\mathbb{P} \in \mathcal{C}_{t_0}(\mu_1, \mu_2)} H(\mathbb{P} | \mathbb{P}_0) \\ &= \sup_{u_1, u_2 \in Lip} \left\{ -\mathbf{1}_{[0, T_1]}(t_0) \int u_1 d\mu_1 - \int u_2 d\mu_2 + u(t_0, X_{t_0}, Y_{t_0}) \right\}, \end{aligned} \quad (4.1.12)$$

where u in the supremum is a viscosity solution of the Hamilton-Jacobi equation

$$\begin{cases} -\partial_t u - \mathcal{L}_{x,y}^0 u + \frac{1}{2} \tau_2^2(x, y) (\partial_y u)^2 = \delta_{T_1}(t) u_1(x), & t \in [0, T_2], \\ u(T_2, x, y) = u_2(x), \end{cases} \quad (4.1.13)$$

where $\mathcal{L}^0$ is as in (4.1.11). Furthermore, if the maximization problem (4.1.12) admits an optimizer u_1^*, u_2^* , then the optimal $\mathbb{P}^*$ in (4.1.12) is given by

$$\left. \frac{d\mathbb{P}^*}{d\mathbb{P}_0} \right|_{\mathcal{F}_{T_2}} = e^{-\int_{t_0}^{T_2} \tau_2(X_t, Y_t) \partial_y u^*(t, X_t, Y_t) dW_t^\perp - \frac{1}{2} \int_{t_0}^{T_2} \tau_2^2(X_t, Y_t) |\partial_y u^*(t, X_t, Y_t)|^2 dt}, \quad (4.1.14)$$

where u^* the solution of (4.1.13) corresponding to u_1^*, u_2^* .

Finally, our method allows us to write a rigorous proof of a weaker version of the formal duality formula that appeared in [Guy22, p.6] and [HL19, Theorem 4.6] and solves the joint SPX-VIX calibration problem. In particular, we consider a function $G : [0, +\infty) \rightarrow \mathbb{R}$ and for $\mathbb{P}$ a probability measure over $(\Omega, \mathcal{F}, \mathbb{P})$ the random variable

$$V'_\mathbb{P} = \mathbb{E}^\mathbb{P} \left[G(X_{T_2}) - G(X_{T_1}) \middle| \mathcal{F}_{T_1} \right]. \quad (4.1.15)$$

and the VIX(G) constraint $\mathcal{L}^\mathbb{P}(V'_\mathbb{P}) \leq_c \mu_3$. We prove the following theorem for (4.1.7) with VIX(G) instead of (4.1.8).

Theorem 4.1.4. *Assume (A3). In addition, we suppose that $|Db| + |D\tau_2| \leq \lambda_0 |\tau_2|$ for some $\lambda_0 > 0$ and that σ is bounded with $\sigma(0, y) = 0$ for every $y \in \mathbb{R}$. Suppose that (X_t, Y_t) has the dynamics (4.1.1) with initial condition $X_{t_0} > 0$ and $Y_{t_0} > 0$ and that μ_1, μ_2, μ_3 are three probability distributions over $\mathbb{R}$ such that $\mu_1 \leq_c \mu_2$ and $\int \psi(x) d\mu_2 < \infty$ for some $\psi : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ with superlinear growth. We further assume that G is bounded, convex and Lipschitz over $[0, +\infty)$ and that (4.1.7) with VIX(G) constraint is admissible. Let u be the*

unique viscosity solution of

$$\begin{cases} -\partial_t u - \mathcal{L}_{x,y}^0 u + \frac{1}{2} \tau_2(x,y)^2 (\partial_y u)^2 = 0, \\ \delta \in \mathbb{R}, t \in (T_1, T_2], (x,y) \in [0, +\infty) \times \mathbb{R}, \\ u(T_2, x, y) = u_2(x) + \delta G(x), & (x,y) \in [0, +\infty) \times \mathbb{R}, \\ u(t, 0, y) = u_2(0) + \delta G(0), & t \in [T_1, T_2], y \in \mathbb{R}. \end{cases} \quad (4.1.16)$$

Let $\Phi(x, y) = \inf_{v \geq 0} \sup_{\delta \in \mathbb{R}} \{u_3(v) - \delta(G(x) + v) + u(T_1^+, x, y; \delta)\}$. Then, Φ is bounded and we have for $t_0 < T_1$

$$\begin{aligned} & \inf_{\mathcal{C}_{t_0}(\mu_1, \mu_2), VIX(G)} (\mathbb{P} | \mathbb{P}_0) \\ &= \sup_{u_1, u_2, u_3} \left\{ - \int u_1 d\mu_1 - \int u_2 d\mu_2 - \int u_3 d\mu_3 + u(t_0, X_{t_0}, Y_{t_0}) \right\}, \end{aligned} \quad (4.1.17)$$

where the supremum is taken over all bounded and Lipschitz u_1, u_2, u_3 is convex and lower bounded, and where $u(t, x, y), t \in [t_0, T_1]$ is a viscosity solution of

$$\begin{cases} -\partial_t u - \mathcal{L}_{x,y}^0 u + \frac{1}{2} \tau_2^2(x,y) (\partial_y u)^2 = 0, & t \in [t_0, T_1], (x,y) \in [0, +\infty) \times \mathbb{R}, \\ u(T_1, x, y) = u_1(x) + \Phi(x, y), & (x,y) \in [0, +\infty) \times \mathbb{R}, \\ u(t, 0, y) = u_1(0) + \Phi(0, y), & t \in [t_0, T_1], y \in \mathbb{R}. \end{cases} \quad (4.1.18)$$

Here, $\mathcal{L}_{x,y}^0$ is given in (4.1.11).

Remark 4.1.5. (i) We note that, due to the dynamics (4.1.1), the assumption that $\sigma(0, y)$ for any $y \in \mathbb{R}$ and the initial condition $X_{t_0} > 0$, we have $X_t \geq 0$ for any $t \geq 0$. This justifies the formula (4.1.15) and the domain in equations (4.1.16), (4.1.18).

(ii) In [Guy22, p.6], the above duality formula (4.1.17), was written formally in the case where $\sigma(x, y) = xy$, $\tau_1(x, y) = \rho \sigma'(y)$ and $\tau_2(x, y) = \sqrt{1 - \rho^2} \sigma'(y)$ for a given $\sigma' : \mathbb{R} \rightarrow \mathbb{R}$.

In addition, instead of the $\text{VIX}(G)$ constraint, which is a constraint on the convex order, the stronger VIX (4.1.8) was considered. Our Theorem 4.1.4 gives a proof of this duality formula when G is $G_n(x) = \max\{-\log(x + \frac{1}{n}), -n\}$, $n \in \mathbb{N}$ and $\text{VIX}(G)$ hold. Since $G_n \xrightarrow{n \rightarrow \infty} G$, our result is an approximation of the formal duality statement [Guy22, p.6].

4.1.5 Organization of the chapter

In section 2, we give the notation, the assumptions we will be using and some preliminary results that will be useful in the proofs. In section 3, we give a quick overview of the (relevant) theory of viscosity solutions for Hamilton-Jacobi equations with the presence of Dirac delta terms. The proofs of Theorem 4.1.1 and Corollary 4.1.2 are given in section 4. In section 5, we prove Theorems 4.1.3 and 4.1.4. In the last section we state some important results from the literature used in the proofs.

4.2 Notation, Assumptions and Preliminary results

4.2.1 Notation

$d \in \mathbb{N}$ is a given dimension, $T, T_1, T_2 > 0$ with $T_1 < T_2$. $\mathbb{R}_{\geq 0} = [0, +\infty)$. γ^d is the d -dimensional standard normal distribution in $\mathbb{R}^d$. For $k \geq 0$, $C_b^k(\mathbb{R}^d)$ is the space of all k -continuously differentiable functions with bounded derivatives. We write $g \in \text{Lip}$ if g is Lipschitz and bounded function. For a function $u = u(t, x) : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}$ depending on time t and space x , we use the symbol Du for the partial derivative $D_x u$.

We use the symbol $\mathcal{P}(\mathbb{R}) = \mathcal{P}_2(\mathbb{R})$ for all probability measures with finite second moments. We denote by $\mathbf{d}_2(\mu, \nu) = \inf_{X \sim \mu, Y \sim \nu} (\mathbb{E}[|X - Y|^2])^{1/2}$ the Wasserstein 2-distance between two measures $\mu, \nu \in \mathcal{P}(\mathbb{R})$.

We say that two probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R})$ are in convex order and we write $\mu \leq_c \nu$.

if $\int h(x)d\mu(x) \leq \int h(x)d\nu(x)$ for every convex function h .⁷ The relative entropy between two measures μ, ν is defined as follows.

$$H(\mu|\nu) = \begin{cases} \int \log \frac{d\mu}{d\nu} d\mu, & \text{if } \mu \ll \nu, \\ +\infty, & \text{otherwise.} \end{cases}$$

When X is a random variable over a probability space $(\Omega, \mathbb{P})$ with distribution μ we write $X \stackrel{\mathbb{P}}{\sim} \mu$ or $\mathcal{L}^{\mathbb{P}}$ or simply $X \sim \mu$, $\mathcal{L}(X) = \mu$, respectively, when there is no confusion about $\mathbb{P}$. If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, we sometimes abuse the notation and we write $V \in \mathcal{F}$ if V is an $\mathcal{F}$ -measurable function. We also use the notation $\mathbb{E}^{\mathbb{P}}$ for the expectation with respect to the probability measure $\mathbb{P}$.

$\mathbf{1}_A$ is the indicator function of a given set A .

4.2.2 Assumptions

We make the following assumptions.

(A1) $L : \mathbb{R} \rightarrow \mathbb{R}$ is convex and p -coercive for some $p \in [2, +\infty)$. That is, there exists a constant C such that $L(a) \geq C|a|^p$ for every $a \in \mathbb{R}$.

(A2) $H : \mathbb{R} \rightarrow \mathbb{R}$ with $H(a) = \sup_{b \geq 0} \{-ab - L(b)\}$ is C^1 .

(A3) $\sigma, b, \tau_1, \tau_2 : \mathbb{R}^2 \rightarrow \mathbb{R}$ are C^2 and Lipschitz functions. τ_2 satisfies $|Db| + |D\tau_2| \leq \lambda_0|\tau_2|$ for some constant $\lambda_0 > 0$.

Remark 4.2.1. (i) Assumption (A1) allows us to prove the duality formula (4.1.9) by providing the compactness needed in order to apply the Von-Neumann theorem (Theorem 4.6.1).

(ii) It is clear that H is convex and degenerate elliptic; that is its derivative $H'(a) \leq 0$ for any $a \in \mathbb{R}$. This allows us to use the theory of viscosity solution for the equation (4.1.10)

7. It is straightforward to show that if m satisfies $\partial_t m = \frac{1}{2} \partial_{xx}(mb)$, for some $b : [0, T] \times \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ then $t \mapsto m_t$ is increasing with respect to the convex order.

(see section 4.3). Assumption (A2) is necessary for the statement of Corollary 4.1.2.

(iii) Assumption (A3) establishes the well-posedness of the stochastic system (4.1.1). Moreover, the C^2 regularity assumption gives us the regularity of the solution of (4.1.13) (Corollary 4.3.5) and (4.1.16) (Proposition 4.3.4).

4.2.3 Preliminary results

In the proofs of our main results, we will be using the following result, which is a compactness theorem related to the Fokker-Planck equation:

$$\partial_t m = \frac{1}{2} \partial_{xx} (b_0 m), \quad (4.2.1)$$

where $b_0 : [t_0, T] \times \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ such that $b_0 \in L^2(dm_s \otimes ds)$. This result is analogous to the one in [BC18, Dau23] which is for Fokker-Planck equations of the form $\partial_t m_k - \Delta m_k + \operatorname{div}(\alpha_k m) = 0$ for some $\alpha_k \in L^2(m_t \otimes dt)$, $k \in \mathbb{N}$, when we have a condition $\int_{t_0}^T \int_{\mathbb{R}^d} |\alpha_k|^2 dm_k(t)(x) dt \leq C$.

Proposition 4.2.2. *Suppose that (A1) holds. Assume that, for all $k \geq 1$, (m_k, b_k) solves the Fokker-Planck equation (3.6.2) starting from m_0 at time t_0 and satisfies the uniform energy estimate*

$$\int_{t_0}^T \int_{\mathbb{R}} L(b_k(t, x)) dm_k(t)(x) dt \leq C, \quad (4.2.2)$$

for some constant $C > 0$ independent of k . Then, there exists $\delta \in (0, 1)$ such that, up to taking a subsequence, $(m_k, b_k m_k)$ converges in $\mathcal{C}^\delta([t_0, T]; \mathcal{P}_2(\mathbb{R})) \times \mathcal{M}([t_0, T] \times \mathbb{R}, \mathbb{R})$ to some (m, w) . The curve m is in $\mathcal{C}^\delta([t_0, T], \mathcal{P}_2(\mathbb{R}))$, w is absolutely continuous with respect to $m(t) \otimes dt$, it holds that

$$\int_{t_0}^T \int_{\mathbb{R}} L\left(\frac{dw}{dm(t) \otimes dt}(t, x)\right) dm(t)(x) dt \leq \liminf_{k \rightarrow +\infty} \int_{t_0}^T \int_{\mathbb{R}} L(b_k(t, x)) dm_k(t)(x) dt \quad (4.2.3)$$

and $(m, \frac{dw}{dm(t) \otimes dt})$ solves (3.6.2) starting from m_0 .

Proof. Note that since b_k is non-negative, we may write $b_k = \sigma_k^2$, where σ_k is the square root of b_k . Then, (3.6.2) describes the law of the process $dX_t = \sigma_k(t, X_t)dB_t$. For any $s, t \in [t_0, T]$, by Itô's formula we have

$$\begin{aligned} \mathbf{d}_2(m_s, m_t) &\leq \mathbb{E}[|X_s - X_t|^2]^{1/2} = \left(\int_s^t \mathbb{E}[\sigma_k(\tau, X_\tau)^2] d\tau \right)^{1/2} \\ &\leq (s - t)^\delta \left(\int_{t_0}^T \int L(b_k(\tau, x)) dm_k(\tau)(x) d\tau \right)^{1-\delta} \leq C(s - t)^\delta, \end{aligned}$$

for some constant $0 < \delta < 1$, because of the assumptions on L and Hölder's inequality. Also, due to (3.4.25) and the coercivity of L , the sequence $b_k m_k$ has bounded variation. By Banach-Alaoglu and Arzela-Ascoli we deduce the convergence of the pair $(m_k, b_k m_k)$ to some (m, w) as claimed.

By Theorem 2.34 of [AFP00], we discover that w is absolutely continuous with respect to $dm(\tau) \otimes d\tau$ and that (3.4.25) holds. Let b be the Radon-Nikodym derivative of w with respect to $dm(\tau) \otimes d\tau$. It is straightforward to show that (m, wm) is a solution to (3.6.2) starting from m_0 . $\square$

Lemma 4.2.3. *Let $0 < T_1 < T_2$, $(\Omega, \mathcal{F}, \mathbb{P} = (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}_0)$ be a filtered probability space and $(f_n)_{n \in \mathbb{N}} : \Omega \rightarrow [0, +\infty)$ a sequence of probability densities that converges weakly in L^1 to some probability density f . Assume further that $(\mathbb{P}_n)_{n \in \mathbb{N}}$ are probability measures over $(\Omega, \mathcal{F})$ such that $d\mathbb{P}_n = f_n d\mathbb{P}$ and V is a random variable in $(\Omega, \mathcal{F})$. The following statements hold: (i) $(\mathbb{P}_n)_{n \in \mathbb{N}}$ converges weakly to the probability measure $\mathbb{P}$, which is such that $d\mathbb{P} = f d\mathbb{P}_0$, (ii) if V is lower bounded, then $\liminf_n \mathbb{E}^{\mathbb{P}_0}[V f_n] \geq \mathbb{E}^{\mathbb{P}_0}[V f]$, and (iii) if V is $\mathcal{F}_{T_1}$ -measurable and bounded, then $\mathbb{E}^{\mathbb{P}_0}[V f_n | \mathcal{F}_{T_1}] \rightharpoonup \mathbb{E}^{\mathbb{P}_0}[V f | \mathcal{F}_{T_1}]$ weakly in L^1 .*

Proof. For (i), we have for any g bounded and continuous

$$\int_{\Omega} g d\mathbb{P}_n = \int_{\Omega} g f_n d\mathbb{P}_0 \xrightarrow{n \rightarrow \infty} \int_{\Omega} g f d\mathbb{P}_0 = \int_{\Omega} g d\mathbb{P}.$$

For (ii), let $N \in \mathbb{N}$ and $K \in \mathbb{R}$ such that $V \geq K$ almost surely. We observe that

$$\mathbb{E}^{\mathbb{P}_0}[Vf_n] = \mathbb{E}^{\mathbb{P}_0}[(V - K)f_n] + K \geq \mathbb{E}^{\mathbb{P}_0}[\mathbf{1}_{\{V-K \leq N\}}(V - K)f_n] + K.$$

Since $\mathbf{1}_{\{V-K \leq N\}}(V - K)$ is bounded, by sending $n \rightarrow +\infty$ we get

$$\liminf_{n \rightarrow +\infty} \mathbb{E}^{\mathbb{P}_0}[Vf_n] \geq \mathbb{E}^{\mathbb{P}_0}[\mathbf{1}_{\{V-K \leq N\}}(V - K)f] + K.$$

The result follows by letting $N \rightarrow +\infty$.

For (iii), let Z be a bounded $\mathcal{F}_{T_1}$ -measurable random variable. We have

$$\int Z \mathbb{E}^{\mathbb{P}_0}[Vf_n | \mathcal{F}_{T_1}] d\mathbb{P}_0 = \int ZVf_n d\mathbb{P}_0 \xrightarrow{n \rightarrow +\infty} \int ZVf d\mathbb{P}_0 = \int Z \mathbb{E}^{\mathbb{P}_0}[Vf | \mathcal{F}_{T_1}] d\mathbb{P}_0.$$

□

4.3 Hamilton-Jacobi Equations

In this section we introduce the definition and the basic properties of solutions of Hamilton-Jacobi equations with the presence of a Dirac delta function. More specifically, equations of the form

$$\begin{cases} -\partial_t u + \mathcal{H}(t, x, u, Du, D^2u) = \delta_{T_1}(t)u_1(x), & (t, x) \in [0, T_2] \times \mathbb{R}^d, \\ u(T_2, x) = u_2(x), & x \in \mathbb{R}^d, \end{cases} \quad (4.3.1)$$

where $d \in \mathbb{N}$ and $\mathcal{H} : [0, T_2] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{d \times d} \rightarrow \mathbb{R}$.

Definition 4.3.1. Let $u_1, u_2 : \mathbb{R}^d \rightarrow \mathbb{R}$ be Lipschitz functions, $0 < T_1 < T_2$, $t \in [0, T_2]$ and $u : [0, T_2] \times \mathbb{R}^d \rightarrow \mathbb{R}$.

(i) We say that u is a viscosity solution of (4.1.10) if u satisfies

$$\begin{cases} -\partial_t u + \mathcal{H}(t, x, u, Du, D^2 u) = 0, & (t, x) \in (T_1, T_2] \times \mathbb{R}^d \\ u(T_2, x) = u_2(x), & x \in \mathbb{R}^d, \end{cases} \quad \text{and} \quad (4.3.2)$$

$$\begin{cases} -\partial_t u + \mathcal{H}(t, x, u, Du, D^2 u) = 0, & (t, x) \in [0, T_1) \times \mathbb{R}^d, \\ u(T_1, x) = u_1(x) + u(T_1^+, x), & x \in \mathbb{R}^d, \end{cases}$$

in the viscosity sense, where $u(T_1^+, x)$ stands for the solution of the problem on the left at time T_1 .

(ii) We say that u satisfies (4.1.10) in the classical sense or that u is a classical solution of (4.1.10) if $u \in BV([0, T]; C_b^2(\mathbb{R}^d))$ and satisfies the equations in (4.3.2) from part (i) in the classical sense.

(iii) We say that u is a classical supersolution of (4.3.1) if $u \in BV([0, T]; C_b^2(\mathbb{R}^d))$ and the equalities in (4.3.2) are satisfied as inequalities: $\leq$.

Following the theory of viscosity solutions of second order Hamilton-Jacobi equations (see [CIL92]), we know that there exists a unique viscosity solution of (4.3.1) when $\mathcal{H}$ satisfies

- $\mathcal{H} \in C([0, T_2] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}_{\text{sym}}^{d \times d})$.
- $\mathcal{H}(t, x, r, p, X) \leq \mathcal{H}(t, x, s, p, Y)$ whenever $r \leq s$ and $Y \leq X$.
- There exists an increasing function $\omega : [0, +\infty) \rightarrow [0, +\infty)$ such that $\omega(0) = 0$ and

$$\mathcal{H}(t, y, r, \alpha(x - y), Y) - \mathcal{H}(t, x, r, \alpha(x - y), X) \leq \omega(\alpha|x - y|^2 + |x - y|)$$

for any $x, y \in \mathbb{R}^d$, $r \in \mathbb{R}$ and $t \in [0, T_2]$, where $\alpha > 0$ and $X, Y \in S_d$ satisfy

$$-3\alpha \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} \leq \begin{pmatrix} X & 0 \\ 0 & -Y \end{pmatrix} \leq 3\alpha \begin{pmatrix} I & -I \\ -I & I \end{pmatrix}.$$

Remark 4.3.2. (i) It is clear that the Hamiltonian $\mathcal{H}$ in Theorem 4.1.1 satisfies these three conditions. The last one holds because H is degenerate elliptic and the inequality between the matrices implies that $X \leq Y$, therefore we can choose $\omega \equiv 0$.

(ii) The Hamiltonian in Theorem 4.1.3 also satisfies these three conditions by similar arguments.

It can be shown that every viscosity solution of (4.3.1) can be approximated by smooth super-solutions of (4.3.1).

Lemma 4.3.3. *Suppose that $\mathcal{H}$ satisfies the above properties and that it is convex in the last three variables. Let u be a bounded viscosity solution of (4.3.1) in the sense of Definition 4.3.1. Then, there exists a sequence $(u_n)_{n \in \mathbb{N}}$ of smooth functions in $([0, T_1] \times \mathbb{R}^d) \cup ((T_1, T_2] \times \mathbb{R}^d)$ such that $u(T_1, x)$ is smooth, $-\partial_t u_n + \mathcal{H}(t, x, u_n, Du_n, D^2 u_n) \leq 0$ for $t \neq T_1$ and $u_n \xrightarrow{n \rightarrow \infty} u$ pointwise. Furthermore $\sup_{(t,x) \in [0, T_2] \times \mathbb{R}^d} |u_n(t, x)| \leq \sup_{(t,x) \in [0, T_2] \times \mathbb{R}^d} |u(t, x)|$ for every $n \in \mathbb{N}$.*

Proof. For $\varepsilon > 0$, we consider the sup-convolution

$$u_\varepsilon(t, x) = \sup_{(s,y) \in [T_1, T_2] \times \mathbb{R}^d} \left\{ u(s, y) - \frac{1}{2\varepsilon}(|x - y|^2 + |t - s|^2) \right\}. \quad (4.3.3)$$

By [KK15, Chapter 4, Theorem 10], we know that u_ε is twice differentiable almost everywhere, $u_\varepsilon \xrightarrow{\varepsilon \rightarrow 0} u$ locally uniformly and satisfies $-\partial_t u_\varepsilon + \mathcal{H}(t, x, u_\varepsilon, Du_\varepsilon, D^2 u_\varepsilon) \leq 0$ almost everywhere. Let $u_\varepsilon^\varepsilon$ be the mollifications of u_ε against a standard ε -mollifier ϕ_ε . Then, from

the convexity of $\mathcal{H}$ we deduce

$$-\partial_t u_\varepsilon^\varepsilon + \mathcal{H}(t, x, u_\varepsilon^\varepsilon, Du_\varepsilon^\varepsilon, D^2 u_\varepsilon^\varepsilon) \leq 0.$$

It is straightforward to show that $u_\varepsilon^\varepsilon(t, x) \xrightarrow{\varepsilon \rightarrow 0} u(t, x)$ for every $(t, x) \in [T_1^+, T_2] \times \mathbb{R}^d$, hence the sequence $(u_{1/n}^{1/n})_{n \in \mathbb{N}}$ has the desired properties in $[T_1^+, T_2]$. We repeat the same procedure (sup-convolution and mollification) to extend $u_{1/n}^{1/n}$ to $[0, T_1] \times \mathbb{R}^d$. We omit the details. $\square$

4.3.1 A regularity result for a special case

We consider the Hamilton-Jacobi equation (4.3.1) with conditions $u_1, u_2 : \mathbb{R}^d \rightarrow \mathbb{R}$ and with Hamiltonian $\mathcal{H} : [0, T] \times \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \times \mathbb{R}^{d \times d}$

$$\mathcal{H}(t, x, u, p, q) = -\text{tr}(\sigma_0(x)\sigma_0^\top(x)q) + \frac{1}{2}\tau_0^2(x)|p|^2 + b_0(x) \cdot p, \quad (4.3.4)$$

where $\sigma_0 : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$, $\tau_0 : \mathbb{R}^d \rightarrow \mathbb{R}_{\geq 0}$ and $b_0 : \mathbb{R}^d \rightarrow \mathbb{R}^d$ are given functions. The following proposition holds for quasilinear parabolic equations with possibly degenerate second order term.

Proposition 4.3.4. *Assume that $\mathcal{H}$ is as in (4.3.4), $\sigma \in C^{1,1}$ and $\tau_0 \in C^2$ such that $|Db_0| + |D\tau_0| \leq \lambda_0|\tau_0|$ for some constant $\lambda_0 > 0$. Let $T > 0$, $g : \mathbb{R}^d \rightarrow \mathbb{R}$ a bounded Lipschitz function. Then, there exists a unique viscosity solution of*

$$\begin{cases} -\partial_t u + \mathcal{H}(x, Du, D^2 u) = 0, & (t, x) \in [0, T] \times \mathbb{R}^d, \\ u(T, x) = g(x), & x \in \mathbb{R}^d. \end{cases}$$

Furthermore, there exists a constant $C > 0$ such that the bound $\|u(t, \cdot)\|_{L^\infty} \leq C\|g\|_{L^\infty}$ gradient bound $\|Du(t, \cdot)\|_{L^\infty} \leq C(T - t + \|Dg\|_{L^\infty})$ hold.

Proof. We regularize $a := \sigma_0 \sigma_0^\top$ as $a_\varepsilon = a + \varepsilon I$ and g using a standard mollifier ρ_ε , $g_\varepsilon = g * \rho_\varepsilon$. Let u^ε be the solution of the (uniformly) parabolic equation

$$\begin{cases} -\partial_t u - \operatorname{tr}(a_\varepsilon D^2 u^\varepsilon) + \frac{1}{2} \tau_0 |Du^\varepsilon|^2 = 0, & (t, x) \in [0, T] \times \mathbb{R}^d \\ u^\varepsilon(T, x) = g_\varepsilon(x), & x \in \mathbb{R}^d. \end{cases}$$

By Bernstein's method, as in [CSS22, Lemma 2.2], the functions $(u^\varepsilon)_{\varepsilon \in (0,1)}$ are uniformly bounded and uniformly Lipschitz, independently of ε . The stability property of viscosity solutions imply that u^ε converges locally uniformly in $[0, T] \times \mathbb{R}^d$ to a function u which is the unique viscosity solution of (4.3.1) (with $\mathcal{H}$ as in (4.3.4)). The proof is complete. $\square$

As a corollary, we obtain a regularity result for (4.3.1) when $\mathcal{H}$ has the form (4.3.4).

Corollary 4.3.5. *Assume that $\mathcal{H}$ is as in (4.3.4), $\sigma_0 \in C^{1,1}$ and $\tau_0 \in C^2$ is such that $|Db_0| + |D\tau_0| \leq \lambda_0 |\tau_0|$ for some constant $\lambda_0 > 0$. Furthermore, $u_1, u_2 : \mathbb{R}^d \rightarrow \mathbb{R}$ are bounded and Lipschitz. Let u be the unique viscosity solution to equation (4.3.1) when (4.3.4) holds. Then, u is bounded and $u(t, \cdot)$ is Lipschitz uniformly in t .*

Proof. By Proposition (4.3.4) on $(T_1, T_2]$, we obtain that $u(\cdot, x)$ is bounded and uniformly Lipschitz when $t \in (T_1, T_2]$. In particular, this implies that $u_1(x) + u(T_1^+, x)$ is bounded and Lipschitz, therefore by Proposition 4.3.4 once again, $u(t, \cdot)$ is bounded and uniformly Lipschitz when $t \in [0, T_1)$. $\square$

4.4 Martingale Optimal Transport, Proof of Theorem 4.1.1

We start by noting the following non-probabilistic interpretation of (4.1.3).

Proposition 4.4.1. *Under the assumptions of Theorem 4.1.1, we have that the minimization*

problem $V(T_0, \mu_0, \mu_1, \mu_2)$ is admissible. Furthermore, the following equality is true

$$V(T_0, \mu_0, \mu_1, \mu_2) = \inf_{b, m} \left\{ \int_{T_0}^{T_2} \int L(b(s, x)) dm_s(x) ds \right. \quad (4.4.1) \\ \left. \left| \partial_t m = \frac{1}{2} \partial_{xx}(mb), \ m(T_0) = \mu_0, \ m(T_1) = \mu_1, \ m(T_2) = \mu_2 \right\}, \right.$$

where the infimum is taken over all $b : [T_0, T_2] \times \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ with $b \in L^2(dm_s ds)$ and we require the equation for m in (4.4.1) to hold in the weak sense.

Proof. The admissibility follows from Theorem 4.4.2 below. The proof of (4.4.1) uses a mimicking theorem and can be found in [GLW22, Lemma 3.1, Proposition 3.4]. $\square$

To motivate the introduction of the Dirac delta in (4.1.10), we consider that at a fixed time $t \in [0, T_1]$ the distribution is μ_0 and we observe that, at least formally, we have

$$\begin{aligned} V(t, \mu_0, \mu_1, \mu_2) &= \inf_{b, m} \sup_{u_1, u_2} \left\{ \int_t^{T_2} \int L(b(s, x)) dm_s(x) ds + \int u_1 dm_{T_1} \right. \\ &\quad \left. - \int u_1 d\mu_1 + \int u_2 dm_{T_2} - \int u_2 d\mu_2 \right\} \\ &= \sup_{u_1, u_2} \inf_{b, m} \left\{ \int_t^{T_2} \int L(b(s, x)) dm_s(x) ds + \int u_1 dm_{T_1} \right. \\ &\quad \left. + \int u_2 dm_{T_2} - \int u_1 d\mu_1 - \int u_2 d\mu_2 \right\} \\ &= \sup_{u_1, u_2} \left\{ - \int u_1 d\mu_1 - \int u_2 d\mu_2 \right. \\ &\quad \left. + \inf_{b, m} \left\{ \int_t^{T_2} \int L(b(s, x)) dm_s(x) ds + \int u_1 dm_{T_1} + \int u_2 dm_{T_2} \right\} \right\}. \end{aligned}$$

Studying the infimum that is inside the supremum is standard in Mean Field Control theory (e.g [CDLL19]), when instead of the term $\int u_1 dm_{T_1}$ we have $\int_t^{T_2} F(s, m(s)) ds$ for some running cost function $F : [0, T_2] \times \mathcal{P}(\mathbb{R}^d) \rightarrow \mathbb{R}$. In our case, formally, $F(s, m) = \int \delta_{T_1}(s) u_1(x) dm(x)$ with $\frac{\delta F}{\delta m}(s, m, x) = \delta_{T_1}(s) u_1(x)$ and hence the optimizer is expected to be associated with the Hamilton-Jacobi equation (4.1.10). In particular, the optimal control

is $b(s, x) = -H' \left(\frac{D^2 u(s, x)}{2} \right)$, $s \in [t, T_2]$, where u solves (4.1.10).

To study (4.4.1), it is reasonable to start with its simplified version; namely when there is no intermediate constraint $m(T_1) = \mu_1$.

We note that if $T_0 \in (T_1, T_2]$, then (4.4.1) becomes independent of μ_1 , therefore the function $V(T_0, \mu, \mu_1, \mu_2)$ is identical to

$$\begin{aligned} & V_2(T_0, \mu_0, \mu_2) \\ &= \inf_{b, m} \left\{ \int_{T_0}^{T_2} \int L(b(s, x)) dm_s(x) ds \middle| \partial_t m = \frac{1}{2} \partial_{xx}(mb), m(T_0) = \mu_0, m(T_2) = \mu_2 \right\}. \end{aligned} \quad (4.4.2)$$

If $T_0 \in [0, T_1]$, for any b_1, b_2 admissible functions satisfying

$$\begin{cases} \partial_t m_1 = \frac{1}{2} \partial_{xx}(m_1 b_1), \\ m_1(T_0) = \mu_0, m_1(T_1) = \mu_1, \end{cases} \quad \text{and} \quad \begin{cases} \partial_t m_2 = \frac{1}{2} \partial_{xx}(m_2 b_2), \\ m_2(T_1) = \mu_1, m_2(T_2) = \mu_2 \end{cases}, \quad (4.4.3)$$

we write $b(s, x) = 1_{[T_0, T_1]}(s)b_1(s, x) + 1_{(T_1, T_2]}(s)b_2(s, x)$ and $m(s) = 1_{[T_0, T_1]}(s)m_1(s) + 1_{(T_1, T_2]}(s)m_2(s)$. The pair (m, b) , then satisfies in the weak sense

$$\begin{cases} \partial_t m = \frac{1}{2} \partial_{xx}(mb), \\ m(T_0) = \mu_0, m(T_1) = \mu_1, m(T_2) = \mu_2, \end{cases} \quad (4.4.4)$$

and hence it is admissible for (4.4.1). Therefore, depending on the value of T_0 , we can study (4.4.1) as

$$\begin{aligned} V_1(T_0, \mu_0, \mu_1, \mu_2) = \inf_{b, m} \left\{ \int_{T_0}^{T_1} \int L(b(s, x)) dm_s(x) ds + V_2(T_1, \mu_1, \mu_2) \middle| \partial_t m = \frac{1}{2} \partial_{xx}(mb), \right. \\ \left. m(T_0) = \mu_0, m(T_1) = \mu_1 \right\}, \end{aligned}$$

if $T_0 \in [0, T_1]$ and $V(T_0, \mu, \mu_1, \mu_2) = V_2(T_0, \mu, \mu_2)$, where V_2 is given in (4.4.2), if $T_0 \in$

$[T_1, T_2]$. To prove Theorem 4.1.1, we will start by proving its following simplified version.

Theorem 4.4.2. *Assume (A1) and (A2). Let $\mu, \nu \in \mathcal{P}(\mathbb{R})$ be two probability measures in convex order $\mu \leq_c \nu$, $T > 0$, $T_0 \in [0, T]$ and*

$$U(T_0, \mu, \nu) := \inf_{b, m} \left\{ \int_{T_0}^T \int L(b(s, x)) dm_s(x) ds \middle| \partial_t m = \frac{1}{2} \partial_{xx}(mb), m(T_0) = \mu, m(T) = \nu \right\}. \quad (4.4.5)$$

Then, $U(T_0, \mu, \nu)$ is admissible and whenever it is finite the following formula holds

$$U(T_0, \mu, \nu) = \sup_{g \in Lip} \left\{ - \int g(x) d\nu(x) + \int u(T_0, x) d\mu(x) \right\}, \quad (4.4.6)$$

where u is a viscosity solution of the Hamilton-Jacobi-Bellman equation

$$\begin{cases} -\partial_t u + H\left(\frac{D^2 u}{2}\right) = 0, & \text{in } [T_0, T] \times \mathbb{R}, \\ u(T) = g, & \text{in } \mathbb{R}. \end{cases} \quad (4.4.7)$$

Before presenting the proof of Theorem 4.4.2, we will show the following Proposition which is related to the Hamilton-Jacobi-Bellman equation (4.4.7).

Proposition 4.4.3. *Assume that (A1), (A2) hold. Let g be a Lipschitz function and $\mu \in \mathcal{P}(\mathbb{R})$. Then, the function $U_g : [0, T] \times \mathcal{P}(\mathbb{R}) \rightarrow \mathbb{R}$ with*

$$U_g(t, \mu) = \inf_{b, m} \left\{ \int_t^T \int L(b(x, s)) dm_s(x) ds + \int g(x) dm_T(x) \right\},$$

where the infimum is taken over all $b : [t, T] \times \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ and $m \in \mathcal{P}(\mathbb{R})$ such that $\partial_s m = \frac{1}{2} \partial_{xx}(bm)$ in the weak sense and $m(t) = \mu$, takes the form $\int u(t, x) d\mu(x)$, where u is a viscosity solution of (4.4.7).

Proof. Fix $(t, \mu) \in [0, T] \times \mathcal{P}(\mathbb{R})$. We notice that $U_g(t, \mu)$ is a mean field control problem

with controls over the diffusion. By the linearity with respect to m of the functional inside the infimum, we have that

$$U_g(t, \mu) = \int_{\mathbb{R}^d} \tilde{U}_g(t, x) \mu(dx),$$

where $\tilde{U}_g(t, x)$ solves (in the classical sense) the stochastic optimal control problem $\tilde{U}_g(t, x) = U_g(t, \delta_x)$. However, by standard stochastic optimal control theory, $U_g(t, \delta_x) = u(t, x)$, where u solves the Hamilton-Jacobi equation

$$\begin{cases} -\partial_s u + H\left(\frac{1}{2}D^2u\right) = 0, & \text{in } [t, T) \times \mathbb{R}, \\ u(T, x) = g(x), & \text{in } \mathbb{R}, \end{cases} \quad (4.4.8)$$

in the viscosity sense. □

We are now ready to prove Theorem 4.4.2.

Proof. Step 1. We will start by showing that $U(t_0, \mu, \nu)$ is finite or, equivalently, that the minimization problem (4.4.5) is admissible. The proof follows an idea from [BBHK17].

Since $\mu \leq_c \nu$, by Strassen's theorem [Str65], there exists a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}\}_{s \geq 0}, \mathbb{P})$ and a discrete martingale (X, Y) adapted to the filtration $(\mathcal{F}_0, \mathcal{F}_1)$ such that $X \sim \mu$ and $Y \sim \nu$. We consider π to be the law of (X, Y) and we write $\pi(dx, dy) = \pi_x(dy) \otimes \mu(dx)$. For any $x \in \mathbb{R}^d$, by Brenier's theorem [Vil21, Theorem 2.12], there exists a convex function ϕ^x such that $(\nabla \phi^x)_\# \gamma^d = \pi_x$. Now let B be a standard Brownian motion adapted to the filtration $\{\mathcal{F}\}_{s \geq 0}$ and

$$M_s := \mathbb{E} \left[\nabla \phi^X(B_1) | \mathcal{F}_s \right]. \quad (4.4.9)$$

It is obvious that M_t is a continuous martingale and for any bounded continuous function h

$$M_0 = \mathbb{E}[\nabla \phi^X(B_1) | \mathcal{F}_0] = \mathbb{E}[\nabla \phi^X(B_1) | X] = \mathbb{E} \left[\int \nabla \phi^X(y) \gamma^d(dy) | X \right]$$

$$\begin{aligned}
&= \mathbb{E}\left[\int z \pi_X(dz) | X\right] = X, \\
\mathbb{E}[h(M_1)] &= \mathbb{E}[h(\nabla \phi^X(B_1)) | \mathcal{F}_1] = \mathbb{E}[h(\nabla \phi^X(B_1))] \\
&= \int h(y) (\nabla \phi^x)_\# \gamma^d(dy) \mu(dx) = \int h(y) \pi(dx, dy) \\
&= \int h(y) \nu(dy),
\end{aligned}$$

hence $M_0 \sim \mu$ and $M_1 \sim \nu$. By the martingale representation theorem there exists a matrix valued process σ_s with $\int_{t_0}^t |(\sigma_{ij})_s|^2 ds < +\infty$ almost surely for each $t \geq 0$, such that $dM_s = \sigma_s dB_s$. However, by the mimicking theorem [BS13, Corollary 3.7], there exists a measurable matrix valued function $\tilde{\sigma}(s, x) = \mathbb{E}[\sigma_s | M_s = x]$ such that the process solving the stochastic differential equation $d\tilde{M}_s = \tilde{\sigma}(t, \tilde{M}_t) dB_s$ on a possibly different probability space satisfies $\mathcal{L}(\tilde{M}_s) = \mathcal{L}(M_s)$ for every $s \geq 0$. Moreover, the curve $t \mapsto m_t = \mathcal{L}(\tilde{M}_s)$ satisfies (3.6.2) in the weak sense with $b(s, x) = \tilde{\sigma} \tilde{\sigma}^\top(s, x)$, initial condition $m_0 = \mu$ and terminal condition $\mu_1 = \nu$. Therefore, the pair $s \mapsto \tilde{m}_s = m_{\frac{s-t_0}{T-t_0}}$ and $\tilde{\sigma}$ is an admissible candidate for the minimization problem (4.4.5).

Step 2. We will now show (4.4.6). For every admissible b , let W be the measure having density b with respect to $m(t) \otimes dt$. We consider the functional

$$\mathcal{L}(g, (W, m)) = \int_{T_0}^T \int L\left(\frac{W}{m}\right) dm_s ds + \int g(x) dm_T(x) - \int g(x) d\nu(x),$$

where g is a Lipschitz function. We will show that $\mathcal{L}$ satisfies the conditions of the Von-Neumann theorem 4.6.1. Indeed, $\mathcal{L}(g, (W, m))$ is concave with respect to g and convex with respect to (W, m) . We now fix a g_* and let C_* such that $C_* > \sup_g \inf_{W, m} \mathcal{L}(g, (W, m))$. Then, if (W_n, m_n) is a sequence such that $\mathcal{L}(g_*, (W_n, m_n)) \leq C_*$, we discover that

$$\int_{T_0}^T \int L\left(\frac{W_n}{m_n}\right) dm_n(s)(x) ds \leq \tilde{C},$$

for some other constant $\tilde{C}$. Hence, by Proposition 4.2.2, (W_n, m_n) converges up to a subsequence to an admissible (W, m) and, by passing to the limit, such that $\mathcal{L}(g_*, W, m) \leq C_*$. Finally, due to our argument from Proposition 4.2.2 once again, $(W, m) \mapsto \mathcal{L}(g, W, m)$ is lower-semicontinuous for every g , thus Theorem 4.6.1 is applicable. We have

$$\begin{aligned} U(T_0, \mu, \nu) &= \inf_{W, m} \sup_{g \in \text{Lip}} \mathcal{L}(g, W, m) \\ &= \sup_{g \in \text{Lip}} \inf_{W, m} \left\{ \int_{T_0}^T \int L\left(\frac{W}{m}\right) dm_s(x) ds + \int g(x) dm_T(x) - \int g(x) d\nu(x) \right\} \\ &= \sup_{g \in \text{Lip}} \left\{ U_g(T_0, \mu) - \int g(x) d\nu(x) \right\}, \end{aligned}$$

where

$$U_g(T_0, \mu) = \inf_{W, m} \left\{ \int_{T_0}^T \int L\left(\frac{W}{m}\right) dm_s(x) ds + \int g(x) dm_T(x) \right\}, \quad (4.4.10)$$

where (W, m) satisfies $\partial_t m = \frac{1}{2} \sum_{i,j=1}^2 \partial_{ij} W_{ij}$ with $m(T_0) = \mu$. By Proposition 4.4.3, U_g satisfies (4.4.7) and the proof of (4.4.6) is complete. $\square$

We will now use Theorem 4.4.2 to prove Theorem 4.1.1.

Proof of Theorem 4.1.1. For $T_0 \geq T_1$ the result follows from Theorem 4.4.2, so we may assume that $T_0 < T_1$. Arguing as in the proof of Theorem 4.4.2, we can show that there exist b_1, b_2 such that (4.4.3) holds, therefore (4.4.1) is admissible. We can once again apply the Von-Neumann theorem to get

$$\begin{aligned} V(T_0, \mu, \mu_1, \mu_2) &= \sup_{u_1, u_2} \inf_{b, m} \left\{ \int_{T_0}^{T_2} \int L(b(s, x)) dm_s(x) ds + \int u_1 dm_{T_1} + \int u_2 dm_{T_2} \right. \\ &\quad \left. - \int u_1 d\mu_1 - \int u_2 d\mu_2 \right\}. \end{aligned}$$

By minimizing on $(T_1, T_2]$ first and then on $[T_0, T_1)$, we get by Theorem 4.4.2

$$\begin{aligned} V(T_0, \mu, \mu_1, \mu_2) &= \sup_{u_1, u_2} \inf_{b, m} \left\{ \int_{T_0}^{T_1} \int L(b(s, x)) dm_s(x) ds + \int (u_1 + u(T_1^+)) dm_{T_1} \right. \\ &\quad \left. - \int u_1 d\mu_1 - \int u_2 d\mu_2 \right\} \\ &= \sup_{u_1, u_2} \left\{ \int u(T_0, x) d\mu(x) - \int u_1(x) d\mu_1(x) - \int u_2(x) d\mu_2(x) \right\}, \end{aligned}$$

where u on the first and second line is viscosity solution of

$$\begin{cases} -\partial_t u + H\left(\frac{D^2 u}{2}\right) = 0, \\ u(T_2) = u_2 \end{cases} \quad \text{and} \quad \begin{cases} -\partial_t u + H\left(\frac{D^2 u}{2}\right) = 0, \\ u(T_1) = u_1 + u(T_1^+), \end{cases}$$

respectively. We conclude that u satisfies (4.1.10) in the sense of Definition 4.3.1.

Finally, we show that the supremum in (4.1.9) can be also taken over classical super-solutions.

We suppose that $T_0 < T_1$. The case $T_0 > T_1$ can be proved similarly. We let

$$D := \sup_v \left\{ \int v(T_0) d\mu_0 - \int v(T_2) d\mu_2 - \mathbf{1}_{[0, T_1]}(T_0) \int (v(T_1^-) - v(T_1^+)) d\mu_1 \right\}, \quad (4.4.11)$$

where the supremum is taken over all smooth v such that $-\partial_t v + H\left(\frac{D^2 v}{2}\right) \leq 0$ when $t \in [T_0, T_1) \cup (T_1, T_2]$. Note that if u is a classical super-solution of (4.1.10) and m satisfies (3.6.2), then if $t_1 < t_2 < T_1$ or $T_1 < t_1 < t_2$ we have

$$\begin{aligned} \int u(t_2) dm_{t_2} - \int u(t_1) dm_{t_1} &\geq \int_{t_1}^{t_2} \int H\left(\frac{D^2 u}{2}\right) dm_s ds + \int_{t_1}^{t_2} \int \frac{D^2 u}{2} b dm_s ds \\ &\geq - \int_{t_1}^{t_2} \int L(b(s, x)) dm_s ds. \end{aligned} \quad (4.4.12)$$

We use (4.4.12) for $t_2 = T_2$ and $t_1 \downarrow T_1$ and $t_1 = T_0$ and $t_2 \uparrow T_1$, respectively, to get

$$\begin{aligned}\int_{T_1}^{T_2} \int L(b(s, x)) dm_s ds &\geq \int u(T_1^+) \mu_1 - \int u_2 d\mu_2, \\ \int_{T_0}^{T_1} \int L(b(s, x)) dm_s ds &\geq \int u(T_0) d\mu_0 - \int u(T_1^-) d\mu_1.\end{aligned}$$

Adding these two relations yields

$$\begin{aligned}\int_{T_0}^{T_2} \int L(b(s, x)) dm_s ds &\geq \int u(T_0) d\mu_0 - \int u(T_2) d\mu_2 + \int (u(T_1^+) - u(T_1^-)) d\mu_1 \\ &\geq \int u(T_0) d\mu_0 - \int u_2 d\mu_2 + \int u_1 d\mu_1,\end{aligned}$$

hence $V(T_0, \mu, \mu_1, \mu_2) \geq D$.

To prove the opposite inequality, let u be a viscosity solution of (4.1.10). Due to Lemma 4.3.3, there exists a uniformly bounded sequence of smooth super-solutions u_n of (4.1.10) (for possibly different u_1, u_2) such that $u_n \xrightarrow{n \rightarrow \infty} u$. By the bounded convergence theorem, this implies

$$\begin{aligned}\lim_{n \rightarrow \infty} \left(\int u_n(T_0) d\mu_0 - \int u_n(T_2) d\mu_2 - \mathbf{1}_{[0, T_1]}(T_0) \int (u_n(T_1^-) - u_n(T_1^+)) d\mu_1 \right) \\ = \int u(T_0) d\mu_0 - \int u_2 d\mu_2 - \mathbf{1}_{[0, T_1]}(T_0) \int u_1 d\mu_1,\end{aligned}$$

hence

$$D \geq \int u(T_0) d\mu_0 - \int u_2 d\mu_2 - \mathbf{1}_{[0, T_1]}(T_0) \int u_1 d\mu_1,$$

Taking supremum over u and using (4.1.9) we get $D \geq V(T_0, \mu_0, \mu_1, \mu_2)$, which finishes the proof. $\square$

Finally, we prove Corollary 4.1.2.

Proof. We assume that $t_0 < T_1$, the other case being similar. Note that if u satisfies (4.1.10) in the classical sense and m satisfies (3.6.2), then if $t_1 < t_2 < T_1$ or $T_1 < t_1 < t_2$ we have

$$\int u(t_2)dm_{t_2} - \int u(t_1)dm_{t_1} = \int_{t_1}^{t_2} \int H\left(\frac{D^2u}{2}\right) dm_s ds + \int_{t_1}^{t_2} \int \frac{D^2u}{2} b dm_s ds. \quad (4.4.13)$$

We first assume that $b = -H'\left(\frac{D^2u}{2}\right)$. Then, (4.4.13) becomes

$$\int u(t_2)dm_{t_2} - \int u(t_1)dm_{t_1} = - \int_{t_1}^{t_2} \int L(b(s, x)) dm_s ds.$$

Therefore, we deduce

$$\begin{aligned} \int_{T_1}^{T_2} \int L(b(s, x)) dm_s ds &= \int u(T_1^+) d\mu_1 - \int u_2 d\mu_2 \\ \int_{T_0}^{T_1} \int L(b(s, x)) dm_s ds &= \int u(T_0) d\mu_0 - \int u(T_1^-) d\mu_1. \end{aligned}$$

We add the last two relations and since $u(T_1^+) - u(T_1^-) = -u_1$, it follows that

$$\int_{T_0}^{T_2} \int L(b(s, x)) dm_s ds = \int u(T_0) d\mu_0 - \int u_1 d\mu_1 - \int u_2 d\mu_2,$$

which means that the supremum in (4.1.9) is achieved.

For the other direction we assume that the supremum in (4.1.9) is achieved at u_1^*, u_2^* . Then, again by (4.4.13), by using the inequality $H\left(\frac{D^2u}{2}\right) + \frac{D^2u}{2}b \geq -L(b)$ we have

$$\int_{T_1}^{T_2} \int L(b(s, x)) dm_s ds \geq \int u(T_1^+) d\mu_1 - \int u_2^* d\mu_2 \quad (4.4.14)$$

$$\int_{t_0}^{T_1} \int L(b(s, x)) dm_s ds \geq \int u(T_0) d\mu_0 - \int u(T_1^-) d\mu_1. \quad (4.4.15)$$

We add (4.4.14) and (4.4.15) to deduce

$$\int_{T_0}^{T_2} \int L(b(s, x)) dm_s ds \geq \int u(T_0) d\mu_0 - \int u_1^* d\mu_1 - \int u_2^* d\mu_2.$$

However, since the supremum in (4.1.9) is achieved, the last holds as an equality, hence (4.4.14) and (4.4.15) are equalities which implies that $H(\frac{D^2 u}{2}) + \frac{D^2 u}{2} b = -L(b)$. It is now straightforward to show that $b(s, x) = -H' \left(\frac{D^2 u(s, x)}{2} \right)$. $\square$

4.5 Martingale Schrödinger bridges, Theorems 4.1.3, 4.1.4

In this section, we give the proofs of Theorems 4.1.3 and 4.1.4.

Proof of Theorem 4.1.3. We denote $\mathbf{W}$ the 2-dimensional Brownian motion $(W, W^\perp)$. Let $\mathbb{P} \ll \mathbb{P}_0$ be a measure in $\mathcal{C}_{t_0}(\mu_1, \mu_2)$. Then, by Girsanov's theorem there exists an adapted process $\alpha_t = (\alpha_t^1, \alpha_t^2)$ such that, for $t \in [t_0, T_2]$, $\frac{d\mathbb{P}}{d\mathbb{P}_0} \Big|_{\mathcal{F}_t} = \exp \left(\int_{t_0}^t \alpha_s d\mathbf{W}_s - \frac{1}{2} \int_{t_0}^t |\alpha_s|^2 ds \right)$ with $\mathbb{E}^{\mathbb{P}} \left[\int_{t_0}^{T_2} |\alpha_t|^2 dt \right] < \infty$. Since X_t is $\mathbb{P}_0$ -martingale, again by Girsanov the process $X_t - \int_{t_0}^t \alpha_s^1 \sigma(X_s, Y_s) ds$ is a $\mathbb{P}$ -martingale. However, $\mathbb{P} \in \mathcal{C}(\mu_1, \mu_2)$, hence X_t is a $\mathbb{P}$ -martingale, which implies that $\int_{t_0}^t \alpha_s^1 \sigma(X_s, Y_s) ds$ is a $\mathbb{P}$ -martingale. We can now easily deduce that $\alpha^1 = 0$, $\mathbb{P}$ -a.s and, therefore, again by Girsanov's theorem, the dynamics (4.1.1) can be written as

$$\begin{cases} dX_t = \sigma(X_t, Y_t) d\widetilde{W}_t, \\ dY_t = (b(X_t, Y_t) + \tau_2(X_t, Y_t) \alpha_t^2) dt + \tau_1(X_t, Y_t) d\widetilde{W}_t + \tau_2(X_t, Y_t) d\widetilde{W}_t^\perp, \end{cases} \quad (4.5.1)$$

where $\widetilde{W}, \widetilde{W}^\perp$ are independent $\mathbb{P}$ -Brownian motions. We have

$$H(\mathbb{P}|\mathbb{P}_0) = \mathbb{E}^{\mathbb{P}} \left[\int_{t_0}^{T_2} \alpha^2 dW_t^\perp - \frac{1}{2} \int_{t_0}^{T_2} |\alpha_s^2|^2 ds \right] = \frac{1}{2} \int_{t_0}^{T_2} \mathbb{E}^{\mathbb{P}} \left[|\alpha_s^2|^2 \right] ds. \quad (4.5.2)$$

Furthermore, by the mimicking theorem [BS13, Corollary 3.6], there exists a measurable function given by $\alpha_2(t, x, y) := \mathbb{E}^\mathbb{P}[\alpha_t^2 | X_t = x, Y_t = y]$ such that for any $t \in [0, T_2]$, $\mathcal{L}^\mathbb{P}(X_t, Y_t) = m_t := \mathcal{L}^\mathbb{P}(\tilde{X}_t, \tilde{Y}_t)$, where (X_t, Y_t) satisfies (4.5.1) and $(\tilde{X}_t, \tilde{Y}_t)$ satisfies

$$\begin{cases} d\tilde{X}_t = \sigma(\tilde{X}_t, \tilde{Y}_t)d\tilde{W}_t, \\ d\tilde{Y}_t = \left(b(\tilde{X}_t, \tilde{Y}_t) + \tau_2(\tilde{X}_t, \tilde{Y}_t)\alpha_2(t, \tilde{X}_t, \tilde{Y}_t)\right)dt + \tau_1(\tilde{X}_t, \tilde{Y}_t)d\tilde{W}_t + \tau_2(\tilde{X}_t, \tilde{Y}_t)d\tilde{W}_t^\perp. \end{cases} \quad (4.5.3)$$

Thus, the left side of (4.1.12) can be written as

$$\begin{aligned} \inf_{\mathbb{P} \in \mathcal{C}_{t_0}(\mu_1, \mu_2)} H(\mathbb{P} | \mathbb{P}_0) &= \inf_{\substack{\alpha_2 \in L^2(dm_t \otimes dt) \\ X_{T_1} \sim \mu_1, X_{T_2} \sim \mu_2}} \left\{ \frac{1}{2} \int_{t_0}^{T_2} \mathbb{E}^\mathbb{P} \left[|\alpha_2(t, \tilde{X}_t, \tilde{Y}_t)|^2 \right] dt \right\} \\ &= \inf_{\alpha_2 \in L^2(dm_t \otimes dt)} \sup_{u_2, u_2 \in Lip} \left\{ \frac{1}{2} \int_{t_0}^{T_2} \mathbb{E}^\mathbb{P} \left[|\alpha_2(t, \tilde{X}_t, \tilde{Y}_t)|^2 \right] dt + \mathbf{1}_{[0, T_1]}(t_0) \int u_1 dm_{T_1} \right. \\ &\quad \left. + \int u_2 dm_{T_2} - \mathbf{1}_{[0, T_1]}(t_0) \int u_1 d\mu_1 - \int u_2 d\mu_2 \right\}. \quad (4.5.4) \end{aligned}$$

By a similar argument as in the proof of Theorem 4.4.2 (Step 2), we can check that the function

$$\begin{aligned} \mathcal{L}((\alpha_2, m), (u_1, u_2)) &:= \frac{1}{2} \int_{t_0}^{T_2} \mathbb{E}^\mathbb{P} \left[|\alpha_2(t, \tilde{X}_t, \tilde{Y}_t)|^2 \right] dt + \mathbf{1}_{[0, T_1]}(t_0) \int u_1 dm_{T_1} + \int u_2 dm_{T_2} \\ &\quad - \mathbf{1}_{[0, T_1]}(t_0) \int u_1 d\mu_1 - \int u_2 d\mu_2 \end{aligned}$$

satisfies the conditions of the Von-Neumann theorem (Theorem 4.6.1)⁸ and hence the infimum in (4.5.4) is a minimum and we can change the order of the infimum and the supremum

8. The compactness is obtained from [Dau23, Proposition 1.2], instead of Proposition 4.2.2.

to find

$$\begin{aligned} & \inf_{\mathbb{P} \in \mathcal{C}_{t_0}(\mu_1, \mu_2)} H(\mathbb{P} | \mathbb{P}_0) \\ &= \sup_{u_1, u_2 \in Lip} \left\{ u(t_0, X_{t_0}, Y_{t_0}) - \int u_2 d\mu_2 - \mathbf{1}_{[0, T_1]}(t_0) \int u_1 d\mu_1 \right\}, \end{aligned} \quad (4.5.5)$$

where $u(t, x, y)$, $t \in [0, T_2]$ is given by

$$\begin{aligned} u(t, x, y) = \inf_{\alpha_2 \in L^2(dm_t \otimes dt)} & \left\{ \frac{1}{2} \int_t^{T_2} \int |\alpha_2(s, x, y)|^2 dm_s(x, y) ds \right. \\ & \left. + \mathbf{1}_{[0, T_1]}(t) \int u_1 dm_{T_1} + \int u_2 dm_{T_2} \right\}, \end{aligned} \quad (4.5.6)$$

where m_t is the law of the process satisfying (4.5.3), as before. We can now check that u satisfies (4.1.13), which finishes the proof of (4.1.12). Indeed, if $t \in (T_1, T_2]$, then by well known results of stochastic optimal control theory, u satisfies (4.1.13) in the viscosity sense in that interval. If $t \in [0, T_1]$, then we break the integral in (4.5.6) and minimize first over $(T_1, T_2]$ and then over $[t, T_1]$ to get

$$u(t, x, y) = \inf_{\substack{\alpha_2 \in L^2(dm_s \otimes ds) \\ s \in [t, T_1]}} \left\{ \int_t^{T_1} \int \frac{|\alpha_2|^2}{2} dm_s ds + \int (u_1(x) + u(T_1^+, x, y)) dm_{T_1}(x, y) \right\},$$

which, again by standard results of stochastic optimal control theory (here the terminal condition is $u(T_1) = u_1 + u(T_1^+)$), satisfies (4.1.13) on $[0, T_1]$.

Finally, if the supremum in (4.1.12) is achieved, we observe that due to Corollary 4.3.5, the solution $u^*(t, x)$ of (4.1.13) (with $u_1 = u_1^*$ and $u_2 = u_2^*$) is bounded and Lipschitz uniformly in $t \in [0, T_1] \cap (T_1, T_2]$. In particular, $\partial_y u^*(t, x, y)$ exists a.e. In addition, $\partial_y u^*(T_1, x, y) = \partial_y (u_1^*(x) + u^*(T_1^+, x, y)) = \partial_y u^*(T_1^+, x, y)$, hence $\partial_y u^*(t, x, y)$ can be defined for all $t \in [0, T_2]$ in a consistent way. To conclude, we note that the optimal α_2 in (4.5.6) is given by $\alpha_2(t, x, y) = -\tau_2(x, y) \partial_y u^*(t, x, y)$, which, combined with $\alpha^1 \equiv 0$, implies (4.1.14).

□

Remark 4.5.1. (i) If they exist, u_1^*, u_2^* are called *Schrödinger potentials*.

(ii) Assume for simplicity that $t_0 = 0$. If the solution u^* of (4.1.13) is classical, then we can write (4.1.14) as

$$\left. \frac{d\mathbb{P}^*}{d\mathbb{P}_0} \right|_{\mathcal{F}_{T_2}} = e^{u^*(0, X_0, Y_0) - u_1^*(X_{T_1}) - u_2^*(X_{T_2}) - \int_0^{T_2} \Delta_t dW_t}, \quad (4.5.7)$$

where $\Delta_t = -\sigma(X_t, Y_t)\partial_x u^*(t, X_t, Y_t) - \tau_1(X_t, Y_t)\partial_y u^*(t, X_t, Y_t)$, which coincides with formula for the optimizer given in [Guy22] (when there is no VIX constraint). Indeed, let $Z_t = (X_t, Y_t)$. We have by Itô's formula, that whenever $0 \leq t_1 < t_2 < T_1$ or $T_1 < t_1 < t_2 \leq T_2$

$$\begin{aligned} & u^*(t_2, Z_{t_2}) - u^*(t_1, Z_{t_1}) \\ &= \frac{1}{2} \int_{t_1}^{t_2} \tau_2^2 |\partial_y u^*|^2 dt + \int_{t_1}^{t_2} \tau_2 \partial_y u^* dW_t^\perp + \int_{t_1}^{t_2} (\sigma \partial_x u^* + \tau_1 \partial_y u^*) dW_t. \end{aligned} \quad (4.5.8)$$

We set $t_1 = 0, t_2 \rightarrow T_1^-$ and $t_1 \rightarrow T_1^+, t_2 = T_2$ in (4.5.8) and we add to discover

$$\begin{aligned} & -u_2^*(X_{T_2}) - u_1^*(X_{T_1}) + \int_0^{T_2} (\sigma \partial_x u^* + \tau_1 \partial_y u^*) dW_t + u^*(0, Z_0) \\ &= - \int_0^{T_2} \tau_2 \partial_y u^* dW_t^\perp - \frac{1}{2} \int_0^{T_2} \tau_2^2 |\partial_y u^*|^2 dt, \end{aligned}$$

which gives that (4.5.7) and (4.1.14). In this case, $-u_1^*(X_{T_1}) - u_2^*(X_{T_2}) - \int_0^{T_2} \Delta_t dW_t$ is called *Schrödinger portfolio*.

(iii) Problem (4.1.5) is not always admissible, therefore in Theorem 4.1.3 we assume that it is. Indeed, if $\sigma(x, y) := 1$ in (4.1.1), then for any $\mathbb{P} \in \mathcal{C}(\mu_1, \mu_2)$ such that $\mathbb{P} \ll \mathbb{P}_0$, the process $(X_t)_{t \in [0, T_2]}$ is a martingale and the distribution $\mathcal{L}^\mathbb{P}((X_t)_{t \in [0, T_2]})$ is absolutely continuous with respect to the distribution of the Brownian motion; hence, its quadratic variation $\langle X \rangle_t = t$, $\mathbb{P}$ -a.s. By Lévy's characterization of Brownian motion, X_t must be a $\mathbb{P}$ -Brownian motion, so $\mathcal{L}^\mathbb{P}(X_{T_1}) = \mathcal{N}(0, T_2)$. This means that $\mu_1 = \mathcal{N}(0, T_2)$, which cannot

be always the case.

To show Theorem 4.1.4, we start by proving two auxiliary results that will be useful in the proof.

Proposition 4.5.2. *Suppose that b, σ satisfy (A3) and $x, y > 0$. For $\delta \in \mathbb{R}$, we consider the optimal control problem*

$$\begin{aligned} u(T_1, x, y; \delta) \\ = \inf_{\alpha} \left\{ \frac{1}{2} \int_{T_1}^{T_2} \int |\alpha(t, x, y)|^2 dm_t dt + \int u_2(x) dm_{T_2} + \delta \int G(x) dm_{T_2} \right\}, \end{aligned} \quad (4.5.9)$$

where the infimum is taken over all square integrable controls α and $(m_t)_{t \in [T_1, T_2]}$ is the distribution of $(\tilde{X}_t, \tilde{Y}_t)$ satisfying

$$\begin{cases} d\tilde{X}_t = \sigma(\tilde{X}_t, \tilde{Y}_t) d\tilde{W}_t, \\ d\tilde{Y}_t = \left(b(\tilde{X}_t, \tilde{Y}_t) + \tau_2(\tilde{X}_t, \tilde{Y}_t) \alpha(t, \tilde{X}_t, \tilde{Y}_t) \right) dt + \tau_1(\tilde{X}_t, \tilde{Y}_t) d\tilde{W}_t + \tau_2(\tilde{X}_t, \tilde{Y}_t) d\tilde{W}_t^\perp, \\ \tilde{X}_{T_1} = x, \quad \tilde{Y}_{T_1} = y, \end{cases}$$

for $t \in [T_1, T_2]$. Then, the map $\delta \mapsto u(T_1, x, y; \delta)$ is continuous.

Proof. Let $\delta_0 \in \mathbb{R}$. We will show that $\delta \mapsto u(T_1, x, y; \delta)$ is continuous at δ_0 .

Since this map is obtained as the infimum of a family of continuous functions, we deduce that it is upper semicontinuous, hence

$$\limsup_{\delta \rightarrow \delta_0} u(T_1, x, y; \delta) \leq u(T_1, x, y; \delta_0). \quad (4.5.10)$$

We will now show the lower semicontinuity. For $\delta \in \mathbb{R}$, we let α^δ be the optimal control in (4.5.9) and we denote by m_t^δ the distribution of (X_t, Y_t) when $\alpha = \alpha^\delta$.

For $\delta \in \mathbb{R}$, by using α^δ as a control we have

$$u(T_1, x, y; \delta_0) - u(T_1, x, y; \delta) \leq (\delta_0 - \delta) \int G(x) dm_{T_2}^\delta \leq |\delta_0 - \delta| \|G\|_{L^\infty}.$$

By sending $\delta \rightarrow \delta_0$, we deduce

$$\liminf_{\delta \rightarrow \delta_0} u(T_1, x, y; \delta) \geq u(T_1, x, y; \delta_0),$$

which is the desired lower semicontinuity. □

In the following Proposition we show that the minimum in (4.1.7) can be attained.

Proposition 4.5.3. *Under the assumptions of Theorem 4.1.4, there exists a $\mathbb{P}^* \in \mathcal{C}_{t_0}(\mu_1, \mu_2)$ such that $\mathcal{L}^{\mathbb{P}^*}(V_{\mathbb{P}^*}) \leq_c \mu_3$ and $D'_{\mathbb{P}_0} = H(\mathbb{P}^*|\mathbb{P}_0)$.*

Proof. Let $(\mathbb{P}_n)_{n \in \mathbb{N}} \in \mathcal{C}_{t_0}(\mu_1, \mu_2)$ such that $\mathcal{L}^{\mathbb{P}_n}(V_{\mathbb{P}_n}) \leq_c \mu_3$ and $H(\mathbb{P}_n|\mathbb{P}_0) \xrightarrow{n \rightarrow \infty} D'_{\mathbb{P}_0}$. Consider the probability densities $f_n := \frac{d\mathbb{P}_n}{d\mathbb{P}_0}$. Then, $H(\mathbb{P}_n|\mathbb{P}_0) = \int f_n \log(f_n) d\mathbb{P}_0$ is a bounded sequence and hence, by the Dunford-Pettis theorem, there exists a probability density f such that $f_n \rightharpoonup f$ weakly in L^1 . We let $\mathbb{P}^*$ to be the probability measure such that $d\mathbb{P}^* = f d\mathbb{P}_0$. It is straightforward to check that $\mathcal{L}^{\mathbb{P}^*}(X_{T_1}) = \mu_1$ and $\mathcal{L}^{\mathbb{P}^*}(X_{T_2}) = \mu_2$.

To show that X_t is a $\mathbb{P}^*$ -martingale, it suffices to show that $\mathbb{E}^{\mathbb{P}^*}[(X_t - X_s)\varphi] = 0$ for any $T_2 \geq t > s \geq 0$ and $\varphi \in L^\infty(\mathcal{F}_s)$. Since X_t is a $\mathbb{P}_n$ -martingale, we know that

$$\mathbb{E}^{\mathbb{P}_0}[f_n(X_t - X_s)\varphi] = 0. \tag{4.5.11}$$

However, by the assumptions on μ_2 , X_{T_2} is uniformly integrable with respect to $(\mathbb{P}_n)_{n \in \mathbb{N}}$ and the la Vallée-Poussin theorem yields a convex function $\psi : \mathbb{R} \rightarrow \mathbb{R}_{\geq 0}$ with superlinear growth such that $\int \psi(x) d\mu_2(x) < +\infty$. Thus, due to the martingality of X_t , we have that

for any $r \in [0, T]$

$$\int \psi(X_r) d\mathbb{P}_n \leq \int \psi(X_{T_2}) d\mathbb{P}_n = \int \psi(x) d\mu_2(x) \leq \infty$$

and hence X_r is uniformly integrable with respect to the family $(\mathbb{P}_n)_{n \in \mathbb{N}}$. Now Lemma 4.2.3 applies and we get $(X_t - X_s)f_n \rightharpoonup (X_t - X_s)f$ weakly in L^1 . The result follows by sending $n \rightarrow \infty$ in (4.5.11).

To show that $\mathcal{L}^{\mathbb{P}^*}(V_{\mathbb{P}^*}) \leq_c \mu_3$, it suffices to show

$$\int h \left(\mathbb{E}^{\mathbb{P}^*} [G(X_{T_2}) - G(X_{T_1}) | \mathcal{F}_{T_1}] \right) d\mathbb{P}^* \leq \int h(x) d\mu_3(x), \quad (4.5.12)$$

for any convex function $h : \mathbb{R} \rightarrow \mathbb{R}$. We set $V = G(X_{T_2}) - G(X_{T_1})$. By using the standard formula $\mathbb{E}^{\mathbb{P}_n}[V | \mathcal{F}_{T_1}] = \frac{\mathbb{E}^{\mathbb{P}_0}[V f_n | \mathcal{F}_{T_1}]}{\mathbb{E}^{\mathbb{P}_0}[f_n | \mathcal{F}_{T_1}]}$, we have by the properties of $(\mathbb{P}_n)_{n \in \mathbb{N}}$

$$\begin{aligned} \int h(x) d\mu_3(x) &\geq \mathbb{E}^{\mathbb{P}_n} \left[h \left(\mathbb{E}^{\mathbb{P}_n} [V | \mathcal{F}_{T_1}] \right) \right] = \int h \left(\frac{\mathbb{E}^{\mathbb{P}_0}[V f_n | \mathcal{F}_{T_1}]}{\mathbb{E}^{\mathbb{P}_0}[f_n | \mathcal{F}_{T_1}]} \right) \mathbb{E}^{\mathbb{P}_0}[f_n | \mathcal{F}_{T_1}] d\mathbb{P}_0 \\ &:= \Phi \left(\mathbb{E}^{\mathbb{P}_0}[V f_n | \mathcal{F}_{T_1}], \mathbb{E}^{\mathbb{P}_0}[f_n | \mathcal{F}_{T_1}] \right), \end{aligned} \quad (4.5.13)$$

where $\Phi : L^1 \times L^1 \rightarrow \mathbb{R}$ is the function defined as

$$\Phi(a, b) = \int h \left(\frac{a}{b} \right) b d\mathbb{P}_0. \quad (4.5.14)$$

Since h is convex, standard arguments imply that Φ is also convex, therefore it is also weakly lower semi-continuous. We now observe that, since V is bounded, Lemma 4.2.3 gives $\mathbb{E}^{\mathbb{P}_0}[f_n | \mathcal{F}_{T_1}] \rightharpoonup \mathbb{E}^{\mathbb{P}_0}[f | \mathcal{F}_{T_1}]$ and $\mathbb{E}^{\mathbb{P}_0}[V f_n | \mathcal{F}_{T_1}] \rightharpoonup \mathbb{E}^{\mathbb{P}_0}[V f | \mathcal{F}_{T_1}]$ weakly in L^1 . With this in mind, passing to the limit in (4.5.13) yields

$$\int h(x) d\mu_3(x) \geq \Phi \left(\mathbb{E}^{\mathbb{P}_0}[V f | \mathcal{F}_{T_1}], \mathbb{E}^{\mathbb{P}_0}[f | \mathcal{F}_{T_1}] \right),$$

which is (4.5.12). □

We are now ready to show Theorem 4.1.4.

Proof of Theorem 4.1.4. The proof follows the lines of the proof of Theorem 4.1.3 but we have to give special treatment to the terms involving the VIX constraint $\mathcal{L}^{\mathbb{P}}(V'_{\mathbb{P}}) \leq \mu_3$. We will be using the notation $\mathcal{P}$ for the probability measures of $(\Omega, \mathcal{F})$ and $\mathcal{P}' = \{\mathbb{P} \in \mathcal{P} : X_t \text{ is a } \mathbb{P}\text{-martingale}\}$. We have

$$\begin{aligned}
D'_{\mathbb{P}_0} &= \inf_{\mathbb{P} \in \mathcal{C}_{t_0}(\mu_1, \mu_2), \text{VIX}(G)} \sup_{u_1, u_2, u_3} \left\{ H(\mathbb{P}|\mathbb{P}_0) + \mathbb{E}^{\mathbb{P}}[u_2(X_{T_2}) + u_1(X_{T_1}) + u_3(V'_{\mathbb{P}})] \right. \\
&\quad \left. - \sum_{i=1}^3 \int u_i(x) \mu_i(dx) \right\} \\
&= \inf_{\mathbb{P} \in \mathcal{P}', d\mathbb{P} = f d\mathbb{P}_0} \sup_{u_1, u_2, u_3} \left\{ \int f \log(f) d\mathbb{P}_0 + \int u_3 \left(\frac{\mathbb{E}^{\mathbb{P}_0}[Wf|\mathcal{F}_{T_1}]}{\mathbb{E}^{\mathbb{P}_0}[f|\mathcal{F}_{T_1}]} \right) \mathbb{E}^{\mathbb{P}_0}[f|\mathcal{F}_{T_1}] d\mathbb{P}_0 \right. \\
&\quad \left. + \mathbb{E}^{\mathbb{P}_0}[u_2(X_{T_2})f] + \mathbb{E}^{\mathbb{P}_0}[u_1(X_{T_1})f] \right. \\
&\quad \left. - \sum_{i=1}^3 \int u_i(x) \mu_i(dx) \right\}, \quad (4.5.15)
\end{aligned}$$

where $W = G(X_{T_2}) - G(X_{T_1})$ and the supremum is taken over all bounded and continuous u_1 , u_2 , and u_3 convex, continuous and lower bounded. We notice that the functional inside the brackets (which we name $\mathcal{G}$), as a function of f and u_1, u_2, u_3 , is convex in $f \in L^1(\Omega)$, linear in u_1, u_2, u_3 and weakly (in L^1) lower semicontinuous with respect to f . In addition, for any fixed u_1, u_2, u_3 and constant C , the set $A = \{f \in L^1 : f \geq 0, \int f d\mathbb{P}_0 = 1, \mathcal{G}(f, u_1, u_2, u_3) \leq C\}$ is compact with respect to the weak topology of L^1 .

Indeed, suppose that $(f_n) \in A$. Since u_1, u_2, u_3 are lower bounded, $\mathcal{G}(f, u_1, u_2, u_3) \leq C$ gives an upper bound for $\int f_n \log(f_n) d\mathbb{P}_0$ and for $\mathbb{E}^{\mathbb{P}_0}[X_{T_2}^2 f_n]$ (by the SDE satisfied by X_t since σ is bounded). On the one hand, the Dunford-Pettis theorem implies that (f_n) converges weakly in L^1 (up to subsequence) to some $f \in L^1(\Omega)$. On the other hand, due to the argument from Proposition 4.5.3, the measure $\mathbb{P}$ with $d\mathbb{P} = f d\mathbb{P}_0$ remains in $\mathcal{C}_{t_0}$. The compactness follows.

The conditions for the Von-Neumann theorem 4.6.1 are satisfied, thus we can write (4.5.15) as

$$D'_{\mathbb{P}_0} = \sup_{u_1, u_2, u_3} \inf_{\mathbb{P} \in \mathcal{P}', d\mathbb{P} = f d\mathbb{P}_0} \left\{ \int f \log(f) d\mathbb{P}_0 + \int u_3 \left(\frac{\mathbb{E}^{\mathbb{P}_0}[Wf|\mathcal{F}_{T_1}]}{\mathbb{E}^{\mathbb{P}_0}[f|\mathcal{F}_{T_1}]} \right) \mathbb{E}^{\mathbb{P}_0}[f|\mathcal{F}_{T_1}] d\mathbb{P}_0 \right. \\ \left. + \mathbb{E}^{\mathbb{P}_0}[u_2(X_{T_2})f] + \mathbb{E}^{\mathbb{P}_0}[u_1(X_{T_1})f] - \sum_{i=1}^3 \int u_i(x) \mu_i(dx) \right\}. \quad (4.5.16)$$

Now, using the fact that $V'_{\mathbb{P}} \geq 0$ whenever $\mathbb{P} \in \mathcal{P}'$ and the equality

$$\mathbb{E}^{\mathbb{P}}[u_3(V'_{\mathbb{P}})] = \inf_{V \in \mathcal{F}_{T_1}} \sup_{\Delta \in L^\infty(\mathcal{F}_{T_1})} \left\{ \mathbb{E}^{\mathbb{P}}[u_3(V) + \Delta(W - V)] \right\},$$

(4.5.16) can be written as

$$D'_{\mathbb{P}_0} = \sup_{u_1, u_2, u_3} \inf_{\mathbb{P} \in \mathcal{P}', d\mathbb{P} = f d\mathbb{P}_0} \left\{ \int f \log(f) d\mathbb{P}_0 + \mathbb{E}^{\mathbb{P}_0}[u_2(X_{T_2})f] + \mathbb{E}^{\mathbb{P}_0}[u_1(X_{T_1})f] \right. \\ \left. - \sum_{i=1}^3 \int u_i(x) \mu_i(dx) \right. \\ \left. + \inf_{V \in \mathcal{F}_{T_1}} \sup_{\Delta \in L^\infty(\mathcal{F}_{T_1})} \left\{ \mathbb{E}^{\mathbb{P}}[u_3(|V|) + \Delta(W - |V|)] \right\} \right\}.$$

With a similar Von-Neumann argument as previously, we can rewrite

$$D'_{\mathbb{P}_0} = \sup_{u_1, u_2, u_3} \inf_{V \in \mathcal{F}_{T_1}} \sup_{\Delta \in L^\infty(\mathcal{F}_{T_1})} \inf_{\mathbb{P} \in \mathcal{P}'} \left\{ H(\mathbb{P}|\mathbb{P}_0) + \mathbb{E}^{\mathbb{P}}[u_2(X_{T_2})] + \mathbb{E}^{\mathbb{P}}[u_1(X_{T_1})] \right. \\ \left. + \mathbb{E}^{\mathbb{P}}[u_3(|V|)] - \sum_{i=1}^3 \int u_i(x) \mu_i(dx) \right. \\ \left. + \mathbb{E}^{\mathbb{P}}[\Delta(G(X_{T_2}) - G(X_{T_1}) - |V|)] \right\}. \quad (4.5.17)$$

By Girsanov's theorem, just as in the proof of Theorem 4.1.3, for every $\mathbb{P} \in \mathcal{P}'$ there exists a process α_t , such that for $t \in [t_0, T_2]$, $\frac{d\mathbb{P}}{d\mathbb{P}_0} \Big|_{\mathcal{F}_t} = \exp \left(\int_{t_0}^t \alpha_s dW_s^\perp - \frac{1}{2} \int_{t_0}^t |\alpha_s|^2 ds \right)$,

$\mathbb{E}^{\mathbb{P}} \left[\int_{t_0}^{T_2} |\alpha_t|^2 dt \right] < \infty$ and (by the mimicking theorem [BS13, Corollary 3.6]) in the minimization problem α can be taken to be a function of t, x, y . Hence, (4.5.17) becomes

$$\begin{aligned} D'_{\mathbb{P}_0} = & \sup_{u_1, u_2, u_3} \inf_{V \in \mathcal{F}_{T_1}} \sup_{\Delta \in L^\infty(\mathcal{F}_{T_1})} \inf_{\alpha \in L^2(dm_t \otimes dt)} \left\{ \frac{1}{2} \int_{t_0}^{T_2} \mathbb{E}^{\mathbb{P}} \left[|\alpha(t, \tilde{X}_t, \tilde{Y}_t)|^2 \right] dt \right. \\ & - \sum_{i=1}^3 \int u_i(x) \mu_i(dx) \\ & + \mathbb{E}^{\mathbb{P}} [u_1(\tilde{X}_{T_1}) + u_2(\tilde{X}_{T_2}) + u_3(|V|)] \\ & \left. + \mathbb{E}^{\mathbb{P}} \left[\Delta \left(G(\tilde{X}_{T_2}) - G(\tilde{X}_{T_1}) - |V| \right) \right] \right\}, \end{aligned} \quad (4.5.18)$$

where m_t is the distribution of $(\tilde{X}_t, \tilde{Y}_t)$ which has the dynamics (under $\mathbb{P}$)

$$\begin{cases} d\tilde{X}_t = \sigma(\tilde{X}_t, \tilde{Y}_t) d\tilde{W}_t, \\ d\tilde{Y}_t = \left(b(\tilde{X}_t, \tilde{Y}_t) + \tau_2(\tilde{X}_t, \tilde{Y}_t) \alpha(t, \tilde{X}_t, \tilde{Y}_t) \right) dt + \tau_1(\tilde{X}_t, \tilde{Y}_t) d\tilde{W}_t + \tau_2(\tilde{X}_t, \tilde{Y}_t) d\tilde{W}_t^\perp, \end{cases} \quad (4.5.19)$$

for independent $\mathbb{P}$ -Brownian motions $\tilde{W}, \tilde{W}^\perp$. By optimizing in $[T_1, T_2]$ first and then using the minmax principle once again, (4.5.18) is equal to

$$\begin{aligned} D'_{\mathbb{P}_0} = & \sup_{u_1, u_2, u_3} \inf_{\alpha \in L^2(dm_t \otimes dt)} \inf_{V \in \mathcal{F}_{T_1}} \sup_{\Delta \in L^\infty(\mathcal{F}_{T_1})} \left\{ \frac{1}{2} \int_{t_0}^{T_1} \int |\alpha(t, x, y)|^2 dm_t dt + \int u_1 dm_{T_1} \right. \\ & - \sum_{i=1}^3 \int u_i d\mu_i \\ & \left. - \mathbb{E}^{\mathbb{P}} \left[\Delta \left(G(\tilde{X}_{T_1}) + |V| \right) - u_3(|V|) - u(T_1, \tilde{X}_{T_1}, \tilde{Y}_{T_1}; \Delta) \right] \right\}, \end{aligned} \quad (4.5.20)$$

where u is the value function of the optimal control problem for $s \in [T_1, T_2]$

$$u(s, x, y; \delta) = \inf_{\alpha} \left\{ \frac{1}{2} \int_s^{T_2} \int |\alpha(t, x, y)|^2 dm_t dt + \int u_2(x) dm_{T_2} + \delta \int G(x) dm_{T_2} \right\} \quad (4.5.21)$$

and the minimization is happening over all drifts α such that m is the distribution of the pair

$(\tilde{X}_t, \tilde{Y}_t)$ satisfying (4.5.19). By standard arguments from stochastic optimal control theory, it is known that there is a unique minimizer $\alpha^\delta(t, x, y) := -\partial_y u(t, x, y; \delta)$ for (3.1.6) and $u(t, x, y; \delta)$ satisfies (in the viscosity sense) in $[0, +\infty) \times \mathbb{R}$

$$\begin{cases} -\partial_t u - \mathcal{L}_{x,y}^0 u + \frac{1}{2} \tau_2^2(x, y) (\partial_y u)^2 = 0, & t \in (T_1, T_2), \\ u(T_2, x, y; \delta) = u_2(x) + \delta G(x), \\ u(t, 0, y; \delta) = u_2(0) + \delta G(0), \end{cases} \quad (4.5.22)$$

where $\mathcal{L}_{x,y}^0$ was defined in (4.1.11). The value of $u(t, 0, y; \delta)$ follows from the fact that if X in (4.5.19) starts from 0, then X remains 0.

Returning to (4.5.20) we see that there is the term

$$\inf_{V \in \mathcal{F}_{T_1}} \sup_{\Delta \in L^\infty(\mathcal{F}_{T_1})} \left\{ \mathbb{E}^\mathbb{P} \left[-\Delta \left(G(\tilde{X}_{T_1}) + |V| \right) + u(T_1, \tilde{X}_{T_1}, \tilde{Y}_{T_1}; \Delta) + u_3(|V|) \right] \right\}. \quad (4.5.23)$$

We observe that the function

$$(\omega, \delta) \mapsto -\delta \left(G(\tilde{X}_{T_1}(\omega)) + |V(\omega)| \right) + u(T_1, \tilde{X}_{T_1}(\omega), \tilde{Y}_{T_1}(\omega); \delta) + u_3(|V(\omega)|)$$

takes finite values, is continuous in δ (Proposition 4.5.2) and measurable in ω (because $(x, y) \mapsto u(T_1, x, y; \delta)$ and u_3 are continuous functions), therefore by Proposition 4.6.2 (case (i)), we can rewrite (4.5.23) as

$$\inf_{V \in \mathcal{F}_{T_1}} \mathbb{E}^\mathbb{P} \left[\sup_{\delta \in \mathbb{R}} \left\{ -\delta \left(G(\tilde{X}_{T_1}) + |V| \right) + u(T_1, \tilde{X}_{T_1}, \tilde{Y}_{T_1}; \delta) + u_3(|V|) \right\} \right].$$

Since the supremum of continuous functions is lower semicontinuous, we may apply Propo-

sition 4.6.2 (case (ii)) to further rewrite (4.5.23) as

$$\mathbb{E}^{\mathbb{P}} \left[\inf_{v \geq 0} \sup_{\delta \in \mathbb{R}} \left\{ -\delta \left(G(\tilde{X}_{T_1}) + v \right) + u(T_1, \tilde{X}_{T_1}, \tilde{Y}_{T_1}; \delta) + u_3(v) \right\} \right] =: \mathbb{E}^{\mathbb{P}} [\Phi(\tilde{X}_{T_1}, \tilde{Y}_{T_1})],$$

where $\Phi(x, y) = \inf_{v \geq 0} \sup_{\delta \in \mathbb{R}} \left\{ -\delta (G(x) + v) + u(T_1, x, y; \delta) + u_3(v) \right\}$. Consequently, equation (4.5.20) becomes

$$\begin{aligned} D'_{\mathbb{P}_0} &= \sup_{u_1, u_2, u_3} \inf_{\alpha \in L^2(dm_t \otimes dt)} \left\{ \frac{1}{2} \int_{t_0}^{T_1} \int |\alpha(t, x, y)|^2 dm_t(x, y) dt + \int u_1(x) dm_{T_1}(x, y) \right. \\ &\quad \left. + \mathbb{E}^{\mathbb{P}} [\Phi(\tilde{X}_{T_1}, \tilde{Y}_{T_1})] - \sum_{i=1}^3 \int u_i(x) d\mu_i(x) \right\} \\ &= \sup_{u_1, u_2, u_3} \left\{ u(t_0, X_{t_0}, Y_{t_0}) - \sum_{i=1}^3 \int u_i d\mu_i \right\}, \end{aligned} \quad (4.5.24)$$

where u is the value function of the optimal control problem for $s \in [t_0, T_1]$

$$\begin{aligned} u(s, x, y) &= \inf_{\alpha} \left\{ \int_s^{T_1} \frac{|\alpha(t, x', y')|^2}{2} dm_t(x', y') dt \right. \\ &\quad \left. + \int (u_1(x') + \Phi(x', y')) dm_{T_1}(x', y') \right\}, \end{aligned} \quad (4.5.25)$$

with the infimum taken over all controls α and m_t is the distribution of the pair $(\tilde{X}_t, \tilde{Y}_t)$ which satisfies (4.5.19) with initial condition at $t = t_0$: $(\tilde{X}_{t_0}, \tilde{Y}_{t_0}) = (x, y)$. We now observe that the function $\Phi : [0, +\infty) \times \mathbb{R} \rightarrow \mathbb{R}$ is bounded, because

$$\begin{aligned} \Phi(x, y) &\geq \inf_{v \geq 0} \{u(T_1, x, y; 0) + u_3(v)\}, \text{ which is lower bounded, and} \\ \Phi(x, y) &\leq \inf_{v \geq 0} \sup_{\delta \in \mathbb{R}} \left\{ -\delta(G(x) + v) + \int u_2(x) dm_{T_2}(x, y) + \delta \int G(x) dm_{T_2}(x, y) + u_3(v) \right\} \\ &= \int u_2(x) dm_{T_2}(x, y) + u_3 \left(\int G(x) dm_{T_2}(x, y) - G(x) \right), \end{aligned}$$

where m_{T_2} is the distribution of (X_{T_2}, Y_{T_2}) from (4.5.19) (with $\alpha_t = 0$), which is upper

bounded. Therefore, (4.5.25) is finite and it is known that u satisfies (in the viscosity sense) the Hamilton-Jacobi equation for $(x, y) \in [0, +\infty) \times \mathbb{R}$

$$\begin{cases} -\partial_t u - \mathcal{L}_{x,y}^0 u + \frac{1}{2} \tau_2^2(x, y) (\partial_y u)^2 = 0, & t \in (t_0, T_1), \\ u(T_1, x, y) = u_1(x) + \Phi(x, y), \\ u(t, 0, y) = u_1(0) + u_2(0) + u_3(0), \end{cases} \quad (4.5.26)$$

where the value of u at $x = 0$ can be obtained as in (4.5.22). Since u satisfies (4.5.22) and (4.5.26) in the viscosity sense, we conclude that u satisfies (4.1.18). The conclusion follows from (4.5.24). $\square$

4.6 Important results form the literature

We state the Von-Neumann theorem a proof of which can be found in [OPS19].

Theorem 4.6.1. *Let $\mathbb{A}$ and $\mathbb{B}$ be convex subsets of some vector spaces and suppose that $\mathbb{B}$ is endowed with some Hausdorff topology. Let $\mathcal{L}$ be a function satisfying:*

$$\begin{aligned} a \mapsto \mathcal{L}(a, b) & \text{ is concave in } \mathbb{A} \text{ for every } b \in \mathbb{B}, \\ b \mapsto \mathcal{L}(a, b) & \text{ is convex in } \mathbb{B} \text{ for every } a \in \mathbb{A}. \end{aligned}$$

Suppose also that there exists $a_* \in \mathbb{A}$ and $C_* > \sup_{a \in \mathbb{A}} \inf_{b \in \mathbb{B}} \mathcal{L}(a, b)$ such that:

$$\begin{aligned} \mathbb{B}_* &:= \{b \in \mathbb{B} \mid \mathcal{L}(a_*, b) \leq C_*\} \text{ is not empty and compact in } \mathbb{B}, \\ b \mapsto \mathcal{L}(a, b) & \text{ is lower-semicontinuous in } \mathbb{B}_* \text{ for every } a \in \mathbb{A}. \end{aligned}$$

Then,

$$\min_{b \in \mathbb{B}} \sup_{a \in \mathbb{A}} \mathcal{L}(a, b) = \sup_{a \in \mathbb{A}} \inf_{b \in \mathbb{B}} \mathcal{L}(a, b),$$

where the fact that the infimum is a minimum is part of the theorem.

Proposition 4.6.2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $R = L^\infty$ or L^1 in $(\Omega, \mathcal{F})$. Suppose that $f : \Omega \times \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty, -\infty\}$ is a function that satisfies one of the following set of conditions:*

(i) *f takes values in $\mathbb{R}$ and $f(\omega, x)$ is measurable in ω and continuous in x .*

(ii) *$f(\omega, x)$ is lower semicontinuous in x and there exists a function $g = g(\omega, y, x) : \Omega \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ which is continuous in (x, y) and measurable in ω such that $f(\omega, x) = \sup_{y \in \mathbb{R}} g(\omega, y, x)$.*

Then,

$$\inf_{X \in R} \int_{\Omega} f(\omega, X(\omega)) d\mathbb{P}(\omega) = \int_{\Omega} \inf_{x \in \mathbb{R}} f(\omega, x) d\mathbb{P}(\omega). \quad (4.6.1)$$

In the (i) case, the infimums can also be replaced by supremums.

Proof. We start by assuming (i). In this case, the result follows from [Roc06, Theorem 3A p.185]. To prove (4.6.1) for supremums, we observe that due to [Roc06, Proposition 2C p.174], $-f$ satisfies the conditions of [Roc06, Theorem 3A p.185], therefore

$$\inf_{X \in R} \int_{\Omega} -f(\omega, X(\omega)) d\mathbb{P}(\omega) = \int_{\Omega} \inf_{x \in \mathbb{R}} \{-f(\omega, x)\} d\mathbb{P}(\omega).$$

Now we assume (ii). By [Roc06, Proposition 2C p.174] and [Roc06, Proposition 2R p.180], f satisfies the conditions of [Roc06, Theorem 3A p.185]. The conclusion follows. $\square$

REFERENCES

- [AB03] Olivier Alvarez and Martino Bardi. Singular perturbations of nonlinear degenerate parabolic pdes: a general convergence result. *Archive for rational mechanics and analysis*, 170:17–61, 2003.
- [AFP00] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford university press, 2000.
- [AL98] Mariko Arisawa and P-L Lions. On ergodic stochastic control. *Communications in partial differential equations*, 23(11-12):2187–2217, 1998.
- [BB00] Jean-David Benamou and Yann Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.
- [BB05] A Bensoussan and GL Blankenship. Singular perturbations in stochastic control. *Singular perturbations and asymptotic analysis in control systems*, pages 171–260, 2005.
- [BBHK17] Julio Backhoff, Mathias Beiglböck, Martin Huesmann, and Sigrid Källblad. Martingale benamou–brenier: a probabilistic perspective. *arXiv preprint arXiv:1708.04869*, 2017.
- [BC18] Ariela Briani and Pierre Cardaliaguet. Stable solutions in potential mean field game systems. *Nonlinear Differential Equations and Applications NoDEA*, 25:1–26, 2018.
- [BCE⁺25] Erhan Bayraktar, Hang Cheung, Ibrahim Ekren, Jinniao Qiu, Ho Man Tai, and Xin Zhang. Viscosity solutions of fully second-order hjb equations in the wasserstein space. *arXiv preprint arXiv:2501.01612*, 2025.
- [BDL17] Rainer Buckdahn, Boualem Djehiche, and Jiongmin Li. Mean-field stochastic differential equations and associated pdes. *The Annals of Probability*, 45(2):824–878, 2017.
- [Ber21] Charles Bertucci. On monotone solutions of mean field games master equations. *Séminaire Laurent Schwartz—EDP et applications*, pages 1–13, 2021.
- [Ber23] Charles Bertucci. Stochastic optimal transport and hamilton-jacobi-bellman equations on the set of probability measures. *arXiv preprint arXiv:2306.04283*, 2023.
- [BEZ23] Erhan Bayraktar, Ibrahim Ekren, and Xin Zhang. Comparison of viscosity solutions for a class of second order pdes on the wasserstein space. *arXiv preprint arXiv:2309.05040*, 2023.

- [BK23] Martino Bardi and Hicham Kouhrouh. Singular perturbations in stochastic optimal control with unbounded data. *ESAIM: Control, Optimisation and Calculus of Variations*, 29:52, 2023.
- [BLL24] Charles Bertucci, Jean-Michel Lasry, and Pierre-Louis Lions. On lipschitz solutions of mean field games master equations. *Journal of Functional Analysis*, 287(5):110486, 2024.
- [BLP11] Alain Bensoussan, Jacques-Louis Lions, and George Papanicolaou. *Asymptotic analysis for periodic structures*, volume 374. American Mathematical Soc., 2011.
- [BS73] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.
- [BS13] Gerard Brunick and Steven Shreve. Mimicking an itô process by a solution of a stochastic differential equation. 2013.
- [BS23a] Zachary William Bezemek and Konstantinos Spiliopoulos. Rate of homogenization for fully-coupled mckean–vlasov sdes. *Stochastics and Dynamics*, 23(02):2350013, 2023.
- [BS23b] Zachary William Bezemek and Konstantinos Spiliopoulos. Rate of homogenization for fully-coupled mckean–vlasov sdes. *Stochastics and Dynamics*, 23(02):2350013, 2023.
- [CD⁺18a] René Carmona, François Delarue, et al. *Probabilistic theory of mean field games with applications I*. Springer, 2018.
- [CD⁺18b] René Carmona, François Delarue, et al. *Probabilistic theory of mean field games with applications I-II*. Springer, 2018.
- [CD⁺18c] René Carmona, François Delarue, et al. *Probabilistic theory of mean field games with applications II*. Springer, 2018.
- [CD22] Alekos Cecchin and François Delarue. Weak solutions to the master equation of potential mean field games. *arXiv preprint arXiv:2204.04315*, 2022.
- [CDJS23] Pierre Cardaliaguet, Samuel Daudin, Joe Jackson, and Panagiotis E Souganidis. An algebraic convergence rate for the optimal control of mckean–vlasov dynamics. *SIAM Journal on Control and Optimization*, 61(6):3341–3369, 2023.
- [CDLL19] Pierre Cardaliaguet, François Delarue, Jean-Michel Lasry, and Pierre-Louis Lions. *The master equation and the convergence problem in mean field games:(ams-201)*. Princeton University Press, 2019.
- [CGK⁺23] Andrea Cosso, Fausto Gozzi, Idris Kharroubi, Huyên Pham, and Mauro Rosatolato. Optimal control of path-dependent mckean–vlasov sdes in infinite-dimension. *The Annals of Applied Probability*, 33(4):2863–2918, 2023.

- [CGK⁺24] Andrea Cosso, Fausto Gozzi, Idris Kharroubi, Huy  n Pham, and Mauro Rosestolato. Master bellman equation in the wasserstein space: Uniqueness of viscosity solutions. *Transactions of the American Mathematical Society*, 377(01):31–83, 2024.
- [CIL92] Michael G Crandall, Hitoshi Ishii, and Pierre-Louis Lions. User’s guide to viscosity solutions of second order partial differential equations. *Bulletin of the American mathematical society*, 27(1):1–67, 1992.
- [CKT23] Giovanni Conforti, Richard Clemens Kraaij, and Daniela Tonon. Hamilton–jacobi equations for controlled gradient flows: the comparison principle. *Journal of Functional Analysis*, 284(9):109853, 2023.
- [CLX22] Peng Chen, Jianya Lu, and Lihu Xu. Approximation to stochastic variance reduced gradient langevin dynamics by stochastic delay differential equations. *Applied Mathematics & Optimization*, 85(2):15, 2022.
- [CM18] Dan Crisan and Eamon McMurray. Smoothing properties of mckean–vlasov sdes. *Probability Theory and Related Fields*, 171:97–148, 2018.
- [CM20] Pierre Cardaliaguet and Marco Masoero. Weak kam theory for potential mfg. *Journal of Differential Equations*, 268(7):3255–3298, 2020.
- [CM23] Andrea Cosso and Mattia Martini. On smooth approximations in the wasserstein space. *Electronic Communications in Probability*, 28:1–11, 2023.
- [CP19] Andrea Cosso and Huy  n Pham. Zero-sum stochastic differential games of generalized mckean–vlasov type. *Journal de Math  matiques Pures et Appliqu  es*, 129:180–212, 2019.
- [CS21] Pierre Cardaliaguet and Panagiotis Souganidis. Weak solutions of the master equation for mean field games with no idiosyncratic noise. *arXiv preprint arXiv:2109.14911*, 2021.
- [CSS22] Pierre Cardaliaguet, Benjamin Seeger, and Panagiotis Souganidis. Mean field games with common noise and degenerate idiosyncratic noise. *arXiv preprint arXiv:2207.10209*, 2022.
- [CST22] Jean-Fran  ois Chassagneux, Lukasz Szpruch, and Alvin Tse. Weak quantitative propagation of chaos via differential calculus on the space of measures. *The Annals of Applied Probability*, 32(3):1929–1969, 2022.
- [CTQ25] Hang Cheung, Ho Man Tai, and Jinniao Qiu. Viscosity solutions of a class of second order hamilton–jacobi–bellman equations in the wasserstein space. *Applied Mathematics & Optimization*, 91(1):23, 2025.

- [Dau23] Samuel Daudin. Optimal control of the fokker-planck equation under state constraints in the wasserstein space. *Journal de Mathématiques Pures et Appliquées*, 175:37–75, 2023.
- [DDJ24] Samuel Daudin, François Delarue, and Joe Jackson. On the optimal rate for the convergence problem in mean field control. *Journal of Functional Analysis*, page 110660, 2024.
- [DJS23] Samuel Daudin, Joe Jackson, and Benjamin Seeger. Well-posedness of hamilton-jacobi equations in the wasserstein space: non-convex hamiltonians and common noise. *arXiv preprint arXiv:2312.02324*, 2023.
- [DS24] Samuel Daudin and Benjamin Seeger. A comparison principle for semilinear hamilton-jacobi-bellman equations in the wasserstein space. *Calculus of Variations and Partial Differential Equations*, 63(4):1–36, 2024.
- [Eva89] Lawrence C Evans. The perturbed test function method for viscosity solutions of nonlinear pde. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 111(3-4):359–375, 1989.
- [Eva92] Lawrence C Evans. Periodic homogenisation of certain fully nonlinear partial differential equations. *Proceedings of the Royal Society of Edinburgh Section A: Mathematics*, 120(3-4):245–265, 1992.
- [FG15] Nicolas Fournier and Arnaud Guillin. On the rate of convergence in wasserstein distance of the empirical measure. *Probability theory and related fields*, 162(3):707–738, 2015.
- [GHLT14] Alfred Galichon, Pierre Henry-Labordere, and Nizar Touzi. A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options. 2014.
- [GLOW21] Ivan Guo, Gregoire Loeper, Jan Obloj, and Shiyi Wang. Optimal transport for model calibration. *arXiv preprint arXiv:2107.01978*, 2021.
- [GLOW22] Ivan Guo, Grégoire Loeper, Jan Oblój, and Shiyi Wang. Joint modeling and calibration of spx and vix by optimal transport. *SIAM Journal on Financial Mathematics*, 13(1):1–31, 2022.
- [GLW17] Ivan Guo, Grégoire Loeper, and Shiyi Wang. Local volatility calibration by optimal transport. *arXiv preprint arXiv:1709.08075*, 2017.
- [GLW22] Ivan Guo, Grégoire Loeper, and Shiyi Wang. Calibration of local-stochastic volatility models by optimal transport. *Mathematical Finance*, 32(1):46–77, 2022.

- [GT77] David Gilbarg and Neil S. Trudinger. *Elliptic Partial Differential Equations of Second Order*, volume 224 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, Berlin, 1st edition, 1977.
- [GT19] Wilfrid Gangbo and Adrian Tudorascu. On differentiability in the wasserstein space and well-posedness for hamilton–jacobi equations. *Journal de Mathématiques Pures et Appliquées*, 125:119–174, 2019.
- [Guy20] Julien Guyon. The joint s&p 500/vix smile calibration puzzle solved. *Risk*, April, 2020.
- [Guy22] Julien Guyon. Dispersion-constrained martingale schrödinger bridges: Joint entropic calibration of stochastic volatility models to s&p 500 and vix smiles. *Available at SSRN 4165057*, 2022.
- [Guy24] Julien Guyon. Dispersion-constrained martingale schrödinger problems and the exact joint s&p 500/vix smile calibration puzzle. *Finance and Stochastics*, 28(1):27–79, 2024.
- [HKWS11] Emily Harvey, Vivien Kirk, Martin Wechselberger, and James Sneyd. Multiple timescales, mixed mode oscillations and canards in models of intracellular calcium dynamics. *Journal of nonlinear science*, 21:639–683, 2011.
- [HL19] Pierre Henry-Labordere. From (martingale) schrodinger bridges to a new class of stochastic volatility model. *Available at SSRN 3353270*, 2019.
- [HLS22] Wei Hong, Shihu Li, and Xiaobin Sun. Diffusion approximation for multi-scale mckean-vlasov sdes through different methods. *arXiv preprint arXiv:2206.01928*, 2022.
- [HLS25] Wei Hong, Shihu Li, and Xiaobin Sun. Diffusion approximation for multi-scale mckean-vlasov sdes through different methods. *Journal of Differential Equations*, 414:405–454, 2025.
- [HLX24] Yuewen Hou, Yun Li, and Longjie Xie. Asymptotic limit of fully coupled multi-scale non-linear stochastic system: the non-autonomous approximation method. *arXiv preprint arXiv:2412.13430*, 2024.
- [HT19] Martin Huesmann and Dario Trevisan. A benamou–brenier formulation of martingale optimal transport. 2019.
- [Kan48] Leonid V Kantorovich. On a problem of monge. In *CR (Doklady) Acad. Sci. URSS (NS)*, volume 3, pages 225–226, 1948.
- [KK15] Nikos Katzourakis and Nikos Katzourakis. Mollification of viscosity solutions and semiconvexity. *An Introduction To Viscosity Solutions for Fully Nonlinear PDE with Applications to Calculus of Variations in L^∞* , pages 49–61, 2015.

- [Kum12] K Suresh Kumar. Singular perturbations in stochastic ergodic control problems. *SIAM Journal on Control and Optimization*, 50(6):3203–3223, 2012.
- [Léo13] Christian Léonard. A survey of the schrödinger problem and some of its connections with optimal transport. *arXiv preprint arXiv:1308.0215*, 2013.
- [Lio07] Pierre-Louis Lions. Cours au college de france. Available at www.college-de-france.fr, 2007.
- [LWX22a] Yun Li, Fuke Wu, and Longjie Xie. Poisson equation on wasserstein space and diffusion approximations for mckean-vlasov equation. *arXiv preprint arXiv:2203.12796*, 2022.
- [LWX22b] Yun Li, Fuke Wu, and Longjie Xie. Poisson equation on wasserstein space and diffusion approximations for mckean-vlasov equation. *arXiv preprint arXiv:2203.12796*, 2022.
- [LWX24] Yun Li, Fuke Wu, and Longjie Xie. Poisson equation on wasserstein space and diffusion approximations for multiscale mckean–vlasov equation. *SIAM Journal on Mathematical Analysis*, 56(2):1495–1524, 2024.
- [MHR07] Ethan A Mastny, Eric L Haseltine, and James B Rawlings. Two classes of quasi-steady-state model reductions for stochastic kinetics. *The Journal of chemical physics*, 127(9), 2007.
- [Mon81] Gaspard Monge. Mémoire sur la théorie des déblais et des remblais. *Mem. Math. Phys. Acad. Royale Sci.*, pages 666–704, 1781.
- [MT06] Toshio Mikami and Michèle Thieullen. Duality theorem for the stochastic optimal control problem. *Stochastic processes and their applications*, 116(12):1815–1835, 2006.
- [OPS19] Carlo Orrieri, Alessio Porretta, and Giuseppe Savaré. A variational approach to the mean field planning problem. *Journal of Functional Analysis*, 277(6):1868–1957, 2019.
- [Pas15] Brendan Pass. Multi-marginal optimal transport: theory and applications. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(6):1771–1790, 2015.
- [Por14] Alessio Porretta. On the planning problem for the mean field games system. *Dynamic Games and Applications*, 4:231–256, 2014.
- [PV01a] Étienne Pardoux and Yu Veretennikov. On the poisson equation and diffusion approximation. 1. *The Annals of Probability*, 29(3):1061–1085, 2001.
- [PV01b] Étienne Pardoux and Yu Veretennikov. On the poisson equation and diffusion approximation. i. *The Annals of Probability*, 29(3):1061–1085, 2001.

- [PV03] É. Pardoux and A. Yu. Veretennikov. On poisson equation and diffusion approximation 2. *The Annals of Probability*, 31(3):1166–1192, 2003.
- [PW17] Huy  n Pham and Xiaoli Wei. Dynamic programming for optimal control of stochastic mckean–vlasov dynamics. *SIAM Journal on Control and Optimization*, 55(2):1069–1101, 2017.
- [QW22a] Huijie Qiao and Wanlin Wei. Efficient filtering for multiscale mckean-vlasov stochastic differential equations. *arXiv preprint arXiv:2206.05037*, 2022.
- [QW22b] Huijie Qiao and Wanlin Wei. Efficient filtering for multiscale mckean-vlasov stochastic differential equations. *arXiv preprint arXiv:2206.05037*, 2022.
- [Roc06] R Tyrrell Rockafellar. Integral functionals, normal integrands and measurable selections. In *Nonlinear Operators and the Calculus of Variations: Summer School Held in Bruxelles 8–19 September 1975*, pages 157–207. Springer, 2006.
- [RSX19a] Michael R  ckner, Xiaobin Sun, and Yingchao Xie. Strong convergence order for slow-fast mckean-vlasov stochastic differential equations. *Annales de l’Institut Henri Poincar  , Probabilit  s et Statistiques*, 2019.
- [RSX19b] Michael R  ckner, Xiaobin Sun, and Yingchao Xie. Strong convergence order for slow-fast mckean-vlasov stochastic differential equations. *Annales de l’Institut Henri Poincar  , Probabilit  s et Statistiques*, 2019.
- [RX21a] Michael R  ckner and Longjie Xie. Diffusion approximation for fully coupled stochastic differential equations. *Annals of Probability*, 49(3):1205–1236, 2021.
- [RX21b] Michael R  ckner and Longjie Xie. Diffusion approximation for fully coupled stochastic differential equations. *Annals of Probability*, 49(3):1205–1236, 2021.
- [Saf89] Mikhail V Safonov. On the classical solution of nonlinear elliptic equations of second order. *Mathematics of the USSR-Izvestiya*, 33(3):597, 1989.
- [Sch31] Erwin Schr  dinger. *  ber die umkehrung der naturgesetze*. Verlag der Akademie der Wissenschaften in Kommission bei Walter De Gruyter u . . . , 1931.
- [Sch32] Erwin Schr  dinger. Sur la th  orie relativiste de l’  lectron et l’interpr  tation de la m  canique quantique. In *Annales de l’institut Henri Poincar  *, volume 2, pages 269–310, 1932.
- [Sha48] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.
- [Sio58] Maurice Sion. On general minimax theorems. 1958.
- [Sri08] Sashi Mohan Srivastava. *A course on Borel sets*, volume 180. Springer Science & Business Media, 2008.

- [Str65] Volker Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36(2):423–439, 1965.
- [SY24] H Mete Soner and Qinxin Yan. Viscosity solutions for mckean–vlasov control on a torus. *SIAM Journal on Control and Optimization*, 62(2):903–923, 2024.
- [TT13] Xiaolu Tan and Nizar Touzi. Optimal transportation under controlled stochastic dynamics. *The annals of probability*, pages 3201–3240, 2013.
- [TTZ23] Mehdi Talbi, Nizar Touzi, and Jianfeng Zhang. Viscosity solutions for obstacle problems on wasserstein space. *SIAM Journal on Control and Optimization*, 61(3):1712–1736, 2023.
- [Vil21] Cédric Villani. *Topics in optimal transportation*, volume 58. American Mathematical Soc., 2021.
- [WLVE05] E Weinan, Di Liu, and Eric Vanden-Eijnden. Analysis of multiscale methods for stochastic differential equations. *Communications on Pure and Applied Mathematics*, 58(11):1544–1585, 2005.
- [WRD12] Wei Wang, AJ Roberts, and Jinqiao Duan. Large deviations and approximations for slow–fast stochastic reaction–diffusion equations. *Journal of Differential Equations*, 253(12):3501–3522, 2012.
- [Zit23] Antonios Zitridis. Homogenization of conditional slow-fast mckean-vlasov sdes. *arXiv preprint arXiv:2308.05874*, 2023.