UNIVERSITY OF CALIFORNIA

Los Angeles

Hamilton Jacobi Equations and Variational Problems in Wasserstein Space

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Mathematics

by

Mohit Bansil

2024

ABSTRACT OF THE DISSERTATION

Hamilton Jacobi Equations and Variational Problems in Wasserstein Space

by

Mohit Bansil
Doctor of Philosophy in Mathematics
University of California, Los Angeles, 2024
Professor Christina Kim, Chair

This thesis explores several problems in the calculus of variations especially in the space of measures. Drawing from mean field games (MFG), Hamilton–Jacobi equations (HJE), optimal transport (OT), and quantum mechanics, we explore challenges in existence, uniqueness, and well-posedness in both deterministic and stochastic settings.

The first part of this work examines mean field games and the associated master equation, an infinite-dimensional partial differential equation that links individual decisions with population dynamics. We establish new well-posedness results under monotonicity and convexity conditions, extending previous theories to accommodate broader interaction structures and types of noise (in particular we obtain results in the absence of idiosyncratic noise).

Next, we investigate Hamilton–Jacobi equations in optimal control and classical mechanics, using canonical transformations as a method to achieve global well-posedness in non-convex settings. This approach expands the range of systems that can be studied, with implications for stability and optimal trajectory analysis. This method also extends to the master equation, revealing hidden monotonicity properties that result in new well-posedness theories.

Finally, we explore an extension OT to quantum mechanics by formulating a quantum dynamic transport problem governed by the Schrödinger equation. By establishing a link with the Pauli problem in quantum state reconstruction, this framework could open new avenues for attacking this long-standing problem.

The dissertation of Mohit Bansil is approved.

Dimitri Y. Shlyakhtenko

Stanley J. Osher

Georg Menz

Christina Kim, Committee Chair

University of California, Los Angeles

2024

To my parents . . .

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I would like to thank Wilfrid Gangbo for mentoring me. I am grateful for him sharing his experiences, opinions, and expertise with me.

I would like to extend my deep gratitude to Chris Anderson, Wilfrid Gangbo, Inwon Kim, and Alpár R. Mészáros for their guidance and support throughout my PhD. I am grateful for the time they have dedicated to mentoring me, and for the opportunities they have provided me to grow as a mathematician. This dissertation would not have been possible without their mentorship.

My sincere thanks go to Georg Menz and Marcus Roper, who have been an extraordinary source of encouragement and support.

I would like to thank my family for their support and encouragement and I would like to thank my friends for their camaraderie.

I would like to thank Chenchen Mou for many helpful discussions related to Chapters 2 and 3 of this dissertation.

Several chapters of this dissertation are based on works that have appeared in other places. Chapter 3 is a version of [BMM23b] which is joint with Alpár R. Mészáros and Chenchen Mou. Chapter 4 is a version of [BM24c] which is joint with Alpár R. Mészáros. Chapter 5 is a version of [BM24b] which is joint with Alpár R. Mészáros.

This work was supported in part by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1650604 and by the Air Force Office of Scientific Research under Award No. FA9550-18-1-0502. I would like to acknowledge the partial support of the Heilbronn Institute for Mathematical Research and the UKRI/EPSRC Additional Funding Programme for Mathematical Sciences through the focused research grant "The master equation in Mean Field Games". I am also grateful for the support of UCLA's Dissertation Year Fellowship.

VITA

2020 B.S. (Mathematics), Michigan State University.

2022 C. Phil. in Mathematics, UCLA, Los Angeles, California

PUBLICATIONS

Mohit Bansil and Alpár R. Mészáros, Hidden monotonicity and canonical transformations for mean field games and master equations. arXiv:2403.05426, 2024. Preprint.

Mohit Bansil and Alpár R. Mészáros, On classical solutions and canonical transformations for Hamilton–Jacobi–Bellman equations. arXiv:2403.05412, 2024. Preprint.

Mohit Bansil, Alpár R. Mészáros, and Chenchen Mou, Global Well-Posedness of Displacement Monotone Degenerate Mean Field Games Master Equations. arXiv:2308.16167, 2023. Preprint.

Mohit Bansil and Jun Kitagawa, An optimal transport problem with storage fees. Electron. J. Differential Equations, vol 2023, paper no. 22, 24pp, 2023.

Mohit Bansil and Jun Kitagawa, Quantitative stability in the geometry of semi-discrete optimal transport. Int. Math. Res. Not. IMRN., no. 10, pp. 7354-7389, 2022.

Mohit Bansil, Computational Semi-Discrete Optimal Transport with General Storage Fees. Journal of Mathematical Analysis and Applications, Volume 503, Issue 1, 2021.

Mohit Bansil and Jun Kitagawa, A Newton algorithm for semi-discrete optimal transport with storage fees. SIAM J. Optim., 31(4), pp. 2586-2613, 2021.

Mohit Bansil and Jun Kitagawa, $\mathcal{W}_\infty$-transport with discrete target as a combinatorial matching problem. Arch. Math. (Basel), 117(2), pp. 189-202, 2021.

Mohit Bansil, Two Stage Algorithm for Semi-Discrete Optimal Transport on Disconnected Domains. arXiv:1910.10617, 2019. Preprint.

Mohit Bansil and Yunus E. Zeytuncu, An Analog of the Weyl Law for the Kohn Laplacian on Spheres. Complex Analysis and its Synergies 6:1, doi:10.1007/s40627-019-0038-0, 2020.

John Ahn, Mohit Bansil, Garrett Brown, Emilee Cardin, and Yunus E. Zeytuncu, Spectra of the Kohn Laplacians on Spheres. Involve vol. 12, no. 5, 855-869, 2019.

CHAPTER 1

Introduction

1.1 Background and Motivation

In recent years, game theory has witnessed a remarkable evolution with the emergence of mean field games (MFGs), a framework that addresses the behavior of large populations of interacting rational agents. Mean field games offer a powerful lens through which to study complex systems characterized by the interplay between individual decision-making and collective dynamics. Central to this framework is the master equation, a fundamental equation that governs the evolution of the system's distribution of agents over state and action spaces.

The study of mean field games is motivated by diverse applications ranging from economics and finance to biology and engineering, where understanding the strategic interactions of a large number of agents is crucial. Traditional game theory approaches encounter challenges when applied to large-scale systems due to the complexity of non-linear PDEs that grow with the number of agents. Mean field games offer a promising avenue to overcome these limitations by approximating the behavior of individual agents within the population, thereby enabling the analysis of systems with a large number of interacting components.

Central to the modern framework of mean field games is the master equation, which governs the evolution of the system's distribution of agents over time. The master equation is remarkable in that it enables a single, comprehensive solution to the MFG problem, even with an arbitrary number of agents. However, its infinite-dimensional nature and nonlinearity present significant challenges for theoretical understanding and solution techniques. Advancements in the well-posedness theory of the master equation deepen our theoretical understanding of MFGs and could have practical implications for optimizing and managing large-scale systems with interacting agents.

A key aspect of mean field games lies in the assumptions made regarding the data. In current works, monotonicity assumptions play a crucial role in ensuring the existence and uniqueness of solutions to the master equation. Similarly, noise is often introduced to provide regularity to the master equation, but understanding the master equation in the absence of noise remains a significant challenge.

Alongside MFGs, optimal transport (OT) has evolved as a versatile mathematical tool for problems involving distributional shifts under constraints. OT theory is traditionally concerned with efficiently moving mass from one distribution to another. We explore an extension of optimal transport into quantum mechanics. Here the dynamics are governed by the Schrödinger equation, creating links between classical OT and quantum mechanics. OT with quantum dynamics involves finding a wave function's optimal evolution from an initial to a final distribution, where optimal is with respect to an energy function in momentum space.

A particularly intriguing connection arises between OT with quantum dynamics and the Pauli problem, which seeks to determine whether a quantum state can be uniquely reconstructed from its (marginal) probability distributions in both position and momentum space. By relating the Pauli problem to a OT problem governed by quantum dynamics, we offer a novel approach to quantum state reconstruction.

1.2 Outline of the Thesis

This thesis explores these frameworks, presenting new results in mean field games, Hamilton–Jacobi equations, and optimal transport with quantum dynamics. Each chapter builds on foundational concepts, extending classical results to address challenges in existence, uniqueness, and stability.

1.2.1 Mean Field Games and the Master Equation

Mean field games (MFGs), developed in the seminal works of Lasry and Lions, and Huang, Malhamé, and Caines ([LL07, HMC06]), provide a rigorous way to model strategic decision-making in populations with a large number of interacting agents. Classical game theory, although powerful in analyzing small groups of rational agents, becomes theoretically and computationally intractable as the number of participants increases. MFG theory circumvents this limitation by modeling the asymptotic behavior of games with infinitely many agents. The resulting mathematical framework captures the dynamics of the population distribution, while also incorporating the strategic interactions among agents.

The core equation in mean field game theory is the *master equation* (introduced in [Lio12a]), a infinite-dimensional partial differential equation (PDE) that links individual agents' strategies with the collective behavior of the entire population. This equation, initially formulated by Lions, governs the evolution of the value function in MFGs and allows one to connect finite-player games with their mean-field limits. In the absence of noise, the deterministic master governs MFGs without noise, while stochastic versions contain terms to account for idiosyncratic noise impacting each agent individually or common noise affecting all agents collectively. The well-posedness of the master equation ensuring existence, uniqueness, and stability of solutions is fundamental to the successful application of MFG theory (see [CDLL19, DLR19, DLR20]), but it requires specific assumptions on the problem data.

Classical well-posedness results rely on either *smallness* conditions (such as restrictions on the time horizon or characteristics of the Hamiltonian) or *monotonicity* conditions on the data, with the two most popular of the latter being Lasry–Lions monotonicity and displacement monotonicity. The former comes from viewing the space of measures as a flat linear space, while the latter leverages concepts from optimal transport, viewing the space of measures as a curved space. In this thesis we are concerned primarily with long time solutions so we will impose monotonicity conditions.

The master equation that we consider in this thesis writes as follows. As data, we are given a Hamiltonian $H : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ and a final cost $G : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$. We emphasize that throughout the text we assume that H and G are smooth enough (we detail the specific assumptions in each chapter), and in particular they are defined and finite at any probability measure with finite second moment. Therefore, they will be assumed to be non-local and regularizing in the measure variable. Furthermore, we are given a time horizon $T > 0$ and the intensities of the Brownian idiosyncratic and common noises $\beta, \beta_0 \in \mathbb{R}$, respectively. Then, the master equation, written for the unknown function $V : (0,T) \times \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ reads as

$$\begin{cases} -\partial_t V(t,x,\mu) + H(x,\mu,\partial_x V) - \mathcal{N}V(t,x,\mu) - \frac{\beta^2}{2}\Delta_{ind}V - \frac{\beta_0^2}{2}\Delta_{com}V(t,x,\mu) & = 0, \\ \qquad in\ (0,T)\times\mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d), & \\ V(T,x,\mu) & = G(x,\mu), \\ \qquad in\ \mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d), & \end{cases} \tag{1.1}$$

where

$$\mathcal{N}V(t,x,\mu) = -\int_{\mathbb{R}^d} \partial_\mu V(t,x,\mu,\tilde{x}) \cdot \partial_p H(\tilde{x},\mu,\partial_x V(t,\tilde{x},\mu))d\mu(\tilde{x})$$
$$\Delta_{ind}V = \mathrm{tr}(\partial_{xx}V(t,x,\mu)) + \int_{\mathbb{R}^d} \mathrm{tr}(\partial_{\tilde{x}\mu}V(t,x,\mu,\tilde{x}))d\mu(\tilde{x})$$

and

$$\Delta_{com}V = \mathrm{tr}(\partial_{xx}V(t,x,\mu)) + \int_{\mathbb{R}^d} \mathrm{tr}(\partial_{\tilde{x}\mu}V(t,x,\mu,\tilde{x}))d\mu(\tilde{x}) + 2\int_{\mathbb{R}^d} \mathrm{tr}(\partial_{x\mu}V(t,x,\mu,\tilde{x}))d\mu(\tilde{x}) + \int_{\mathbb{R}^d\times\mathbb{R}^d} \mathrm{tr}(\partial_{\mu\mu}V(t,x,\mu,\tilde{x},\bar{x}))d\mu(\tilde{x})d\mu(\bar{x}).$$

Here $\partial_\mu V$ stands for the so-called Wasserstein gradient whose definition is given later in the text.

Chapter 2 of this thesis addresses the deterministic master equation ($\beta = \beta_0 = 0$). We present a new well-posedness result for the deterministic master equation, particularly for displacement monotone systems, which expand classical solutions to cases lacking regularizing noise effects. Furthermore, we don't require a 'separability' condition on the Hamiltonian, which is a significant departure from previous results (such as [GM22a]).

Chapter 3 extends these results to the setting where we have common noise but no idiosyncratic noise. The chapter discusses the role of displacement monotonicity conditions that are sufficient to guarantee well-posedness, even under degenerate noise conditions. To the best of our knowledge, these results represent one of the first well-posedness theories for long time classical solutions of the master equation in MFGs in the presence of common noise without idiosyncratic noise (some weak solution theory is obtained in [CS22a, CS22b, CSS22]).

1.2.2 Hamilton–Jacobi Equations and Canonical Transformations

Hamilton–Jacobi equations (HJEs) originate in classical mechanics (see [Arn89]), where they encode the system's dynamics via the principle of least action. This principle states that the actual path taken by a system between two states is the one that minimizes (or extremizes) the action, a scalar quantity that summarizes the effect of the system's dynamics over time. HJEs arise naturally in Lagrangian and Hamiltonian mechanics, providing an elegant way to study mechanical problems.

In optimal control theory, the HJE describes the value function representing the minimal cost for a system to transition from one state to another. Dynamic programming principles lead to the Hamilton–Jacobi–Bellman equation, a first-order PDE that governs the evolution of this value function (see [Eva98]). Solutions to the HJE can often be challenging to obtain due to non-linearities and the potential for singularities to form over time. Additionally, the absence of convexity in certain problems can prevent global well-posedness, limiting the applicability of classical solution techniques.

Chapter 4 of this thesis introduces a novel approach to achieving global well-posedness in Hamilton–Jacobi equations through the use of *canonical transformations*. These transformations, which preserve the symplectic structure of Hamiltonian mechanics, enable a redefinition of the Hamiltonian in a way that can reveal convexity properties. Specifically, transformations that are linear in the phase space—mapping $(x, p) \mapsto (x, p - \alpha x)$, for instance—allow one to convert certain Hamiltonians arising from non-convex Lagrangians into a convex-concave form. This reformulation facilitates the analysis of global well-posedness for the system, expanding the set of Hamiltonians that can be studied effectively under this framework.

Chapter 5 extends these results to the master equation of MFG. We leverage the canonical transformations to reveal hidden monotonicity in non-monotone data. By reformulating the Hamiltonian through these transformations, we extend the well-posedness theory of the master equation to cover MFG scenarios that do not satisfy traditional monotonicity assumptions.

1.2.3 Optimal Transport and Quantum Mechanics

Optimal transport theory provides a way to study the efficient movement of distributions under given cost functions (we refer to [Vil03, Vil09] for a thorough introduction). Traditionally, optimal transport is governed by deterministic maps; however, Eric Carlin and Wilfrid Gangbo have extended it to settings where transport follows quantum dynamics, particularly through the Schrödinger equation. This quantum extension of OT links classical transport with quantum mechanics, offering a framework to understand probabilistic distributions in a quantum setting.

Chapter 6 investigates this quantum dynamic extension of optimal transport, connecting it with the Pauli problem, a classical question in quantum mechanics. The Pauli problem asks whether a quantum state can be uniquely determined by its magnitude distributions in both position and momentum space (see [CH78, Cor06] for some partial results). In the

OT with quantum dynamics framework, this translates to minimizing the kinetic energy of a system subject to constraints on initial and final probability densities.

This chapter establishes an equivalence between the Pauli problem and an OT problem governed by quantum dynamics. The results provide a new perspective on quantum state reconstruction, with many potential applications.

1.3 Contributions and Significance

1. **New well-posedness results for mean field game master equations:** This thesis extends classical well-posedness results for the master equation in mean field games, incorporating versions of the well-known displacement monotonicity conditions that ensure stability and uniqueness for both deterministic and stochastic models.

2. **Application of canonical transformations in Hamilton–Jacobi theory:** By applying canonical transformations to the analysis of Hamilton–Jacobi equations, this work expands the class of systems that admit globally well-posed solutions, opening up new applications in optimal control.

3. **Quantum extension of optimal transport and applications to the Pauli problem:** By exploring optimal transport with quantum dynamics, this thesis offers a novel approach to studying the Pauli problem, enriching our understanding of quantum state reconstruction and contributing to the intersection of OT and quantum mechanics.

Together, these results advance the mathematical frameworks for studying complex, high-dimensional systems across regimes, from deterministic and stochastic dynamics to quantum mechanical behaviors.

CHAPTER 2

Deterministic Master Equation

2.1 Introduction and Motivation

2.1.1 Motivation for Mean Field Games

In recent decades, the need to model large populations of interacting agents—ranging from economic agents, such as consumers and firms, to biological populations and self-driving vehicles—has led to the emergence of mean field game (MFG) theory. Classical game theory, while providing powerful tools for modeling strategic interactions among a small number of agents, becomes intractable as the number of participants grows. MFG theory addresses this limitation by studying the asymptotic behavior of games with infinitely many agents, allowing for a continuum approach to model the population's overall dynamics. This framework, initiated independently by Lasry and Lions and by Huang, Caines, and Malhamé, has rapidly developed into a robust mathematical foundation for understanding collective behavior in complex systems.

At the heart of MFG theory is the idea that, in large populations, individual actions have a negligible impact on the overall distribution, allowing each agent to interact with an "average effect" of the population. This leads to a model where each agent optimally responds to the statistical distribution of all agents rather than to specific interactions, dramatically reducing the complexity of analyzing such systems. For instance, MFGs provide insights into traffic flow optimization, where individual drivers' decisions (such as route choice) collectively influence congestion patterns, or into financial markets, where the trading behavior of numerous investors shapes the aggregate demand and price evolution.

MFG theory also enables the quantitative analysis of long-term dynamics and equilibrium properties of large populations. It bridges the gap between microscopic and macroscopic descriptions, offering a rigorous framework to derive macroscopic equations that describe the collective behavior of agents from the underlying microscopic interactions. In economic applications, MFGs allow for the study of competitive behaviors in large markets, such as price formation and product diffusion, which have critical implications for policy design and resource allocation.

Furthermore, MFGs have significant mathematical appeal, as they often involve studying systems of coupled partial differential equations (PDEs) with intricate interactions between the agents' states and distributions. These systems require innovative approaches in PDE theory, probability, and optimal control, making MFGs a fertile ground for theoretical advances. Additionally, MFGs provide a framework for exploring questions of existence, uniqueness, and stability of equilibria in dynamic systems, as well as foundational probabilistic results, such as large deviation principles and central limit theorems for interacting particle systems.

In summary, MFG theory is a versatile and powerful tool for analyzing complex, large-scale systems in which individual decisions collectively shape emergent behaviors. By reducing complex interactions to tractable models that retain essential features of the system, MFGs open up new possibilities for understanding, predicting, and controlling dynamics in vast populations across various scientific fields.

2.1.2 Introduction to the Deterministic Master Equation

The deterministic master equation, a fundamental object in mean field game (MFG) theory, provides a framework for describing the limit behavior of large populations of interacting agents in the absence of random perturbations. Initially introduced by P.-L. Lions in his lectures at Collège de France [Lio12b], this partial differential equation (PDE) governs the evolution of both individual agent states, typically within a finite-dimensional Euclidean space, and the overall distribution of agents, often represented by a Borel probability measure over this space. In a purely deterministic setting, solutions to the master equation play a critical role in analyzing Nash equilibria, capturing the strategic interactions within a population as it approaches infinite size, and offering insight into the stability and uniqueness of these equilibria in the absence of stochastic fluctuations (see [CDLL19]).

The deterministic nature of the master equation poses unique challenges and has thus motivated several specialized approaches within the literature. Ensuring global well-posedness, particularly for classical solutions, requires that the model data satisfy specific conditions. These conditions can be broadly classified into two types: (i) those imposing a smallness criterion on the time horizon, and (ii) those enforcing monotonicity, such as the Lasry–Lions monotonicity condition. In many deterministic settings, the Lasry–Lions monotonicity condition has proven essential for establishing uniqueness of Nash equilibria in the underlying game and, by extension, ensuring well-posedness of the master equation under certain regularity conditions on the data (see [CP20a, CCD22, GM22b]).

Unlike the case with non-degenerate idiosyncratic noise, the deterministic setting is less well understood. To the best of our knowledge the only other long time classical well-posedness result for the deterministic master equation is [GM22a], which requires a potential structure and separable Hamiltonian.

Despite significant advances, deterministic master equations continue to present challenges, particularly in understanding solution behavior when monotonicity conditions are violated. In such cases, standard well-posedness results no longer hold, and weak solution concepts may be necessary. Recent efforts, such as those in [CD24] and [GM22b], propose frameworks to handle these complexities, though the development of robust weak solution theories remains an active area of research.

2.1.3 Motivation for Displacement Monotonicity

Displacement monotonicity is an important concept in the study of mean field games (MFGs), as it provides a robust condition for ensuring the uniqueness and stability of solutions, particularly in large-scale systems. To understand displacement monotonicity in MFGs, it helps to first explore displacement convexity, a foundational idea from optimal transport theory that generalizes classical convexity to probability measures.

2.1.3.1 Displacement Convexity vs. Regular Convexity

Convexity defines a function $f(x)$ as convex if, for all points x and y in its domain and any $\lambda \in [0, 1]$, the function satisfies:

$$f(\gamma_\lambda) \leq \lambda f(x) + (1 - \lambda) f(y).$$

where γ_t is the geodesic connecting x and y. In the case where the domain of f is a Euclidean space the geodesics are straight lines and the above becomes

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y).$$

This inequality indicates that f lies below the line segment connecting $f(x)$ and $f(y)$, which is key in many areas of analysis and optimization. Convex functions have well-known stability and uniqueness properties, which make them easier to optimize and analyze.

Displacement convexity extends this idea to functionals defined on probability measures. Specifically, consider a functional $\mathcal{F}(\mu)$, where μ is a probability measure over some space (often $\mathbb{R}^d$). A functional $\mathcal{F}(\mu)$ is said to be *displacement convex* if, for any two probability measures μ_0 and μ_1, we have:

$$\mathcal{F}(\mu_\lambda) \leq (1 - \lambda)\mathcal{F}(\mu_0) + \lambda \mathcal{F}(\mu_1),$$

where the interpolation, μ_λ, between μ_0 and μ_1 is defined along the *Wasserstein geodesic* the path in probability space that minimizes the transportation cost to move the distribution μ_0

to μ_1. This notion arises in optimal transport theory, where the Wasserstein space is a metric space for probability measures, equipped with the Wasserstein distance that quantifies the "cost" of transporting mass from one distribution to another.

In displacement convexity, the convexity is not along linear paths but rather along geodesic paths in Wasserstein space, which follow the direction of minimal transport cost. This provides a generalized notion of convexity suited for spaces of probability measures, allowing $\mathcal{F}$ to retain stability properties similar to those of classically convex functions in contexts where mass redistributes optimally.

2.1.3.2 Mean Field Games as Optimization Problems

In certain mean field games, referred to as *potential mean field games (MFGs)*, the data (the Hamiltonian and boundary condition) can be described by a potential functional $\mathcal{J}(\mu)$, where μ is a probability measure representing the distribution of the population over states. In these settings, the behavior of the system can be characterized as an optimization problem, where the evolution of the agents' distribution is driven by minimizing a functional, rather than by solving a traditional Nash equilibrium.

When the functional $\mathcal{J}$ is displacement convex, the data can be shown to satisfy a property called *displacement monotonicity*. Displacement convexity in this context implies that $\mathcal{J}$ has a unique minimizer in Wasserstein space, allowing the MFG to reach a stable equilibrium where the distribution μ does not oscillate or create multiple equilibria as agents adjust their strategies.

As we will see below it is possible to formulate this displacement monotonicity condition without needing a potential structure (this was realized in [GMMZ22]). This condition can replace the Lasry–Lions monotonicity condition and in fact is able to give a well-posedness result for the master equation without needing idiosyncratic noise (as we will see below).

2.1.4 From Optimal Control to the Master Equation

2.1.4.1 Optimal Control

Optimal control theory addresses problems where a single agent, beginning at some initial point $y \in \mathbb{R}^d$, aims to follow a trajectory that minimizes an associated cost over time. This control trajectory, represented by $\gamma : [0,T] \to \mathbb{R}^d$, is chosen to balance an ongoing movement cost with a terminal cost upon reaching a destination. The agent's objective is to control its velocity, which influences the trajectory, and to minimize the combined running cost $L(x,v)$ and final cost $G(x)$ at the endpoint. The agent's optimization problem is given by:

$$\inf_{\gamma} \int_0^T L(\gamma(s), \dot{\gamma}(s))\, ds + G(\gamma(T)),$$

where $L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ quantifies the instantaneous cost based on the position x and velocity v, and $G : \mathbb{R}^d \to \mathbb{R}$ determines the penalty or cost at the terminal state $\gamma(T)$.

To solve this, dynamic programming is applied. This method utilizes the fact that any optimal path γ retains optimality over all subintervals within $[0,T]$. The concept of a value function $V(t,x)$ arises, representing the minimal possible cost from state x at time t onward. For any subinterval $[t,T]$, this value function satisfies a partial differential equation called the Hamilton–Jacobi–Bellman (HJB) equation:

$$\partial_t V + \inf_v \left(L(x,v) + \nabla_x V \cdot v \right) = 0.$$

In this equation, the Hamiltonian $H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is defined as:

$$H(x,p) = \sup_v \left(\langle p, v \rangle - L(x,-v) \right),$$

which recasts the HJB equation as:

$$\partial_t V - H(x, \nabla_x V) = 0,$$

with the boundary condition $V(T,x) = G(x)$ at the terminal time T.

2.1.4.2 Differential Games

Differential game theory generalizes optimal control by involving multiple agents or players, each with competing objectives that depend on the actions of others. In a differential game with N players, each player i starts at an initial position $x_i \in \mathbb{R}^d$ and controls their own velocity to minimize their individual cost, while interacting with the strategies of other players.

Each player incurs a running cost $L_i : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^{d(N-1)} \to \mathbb{R}$, based on their own state, velocity, and the states of the other players. They also face a terminal cost $G_i : \mathbb{R}^d \times \mathbb{R}^{d(N-1)} \to \mathbb{R}$, depending on their final position relative to the other players. Player i's total cost for a path γ_i is:

$$\begin{aligned}&\int_0^T L_i(\gamma_i(s), \dot{\gamma}_i(s), \gamma_1(s), \ldots, \gamma_{i-1}(s), \gamma_{i+1}(s), \ldots, \gamma_N(s))\, ds \\ &+ G_i(\gamma_i(T), \gamma_1(T), \ldots, \gamma_{i-1}(T), \gamma_{i+1}(T), \ldots, \gamma_N(T)).\end{aligned}$$

In this framework, each player is assumed to be rational and seeks to optimize their individual cost. This interdependence creates a competitive scenario in which the best response of each player depends on the strategies of others. A solution concept called the Nash equilibrium arises, where no player can unilaterally reduce their cost by changing only their strategy, given the strategies of the others. Mathematically, a Nash equilibrium is a collection of paths $(\gamma_1, \ldots, \gamma_N)$ where each γ_i is optimal for player i, given the paths of all other players.

The dynamic programming approach for differential games leads to a coupled system of PDEs for the players' value functions $V_i(x, t, z)$. Each value function V_i describes the minimal cost for player i from a given state and time, while accounting for the strategies of others:

$$\partial_t V_i(x,t,z) - H_i(x, \nabla_x V_i(x,t,z), z) - \sum_{j \neq i} \nabla_{z_j} V_i(x,t,z) \cdot \nabla_p H_j(z_j, \nabla_x V_j(z_j,t,z), z) = 0,$$

where H_i is the Hamiltonian associated with the cost function L_i of player i. This equation illustrates the intricate interdependencies between players' strategies, creating challenges in establishing well-posedness and uniqueness of Nash equilibria, particularly when the players' objectives conflict or align in complex ways.

2.1.4.3 The Master Equation for Deterministic Mean Field Games

In mean field games (MFGs), when the number of agents N becomes large, interactions between any two individual agents become negligible. Each agent optimally responds to the distribution of the overall population rather than to specific agents, making MFGs particularly suitable for large-scale systems. In the deterministic setting, we assume the absence of random perturbations (or noise) affecting individual agents. Consequently, the collective behavior is governed by a master equation, which describes the evolution of the value function V over time and space.

Assume all agents are identical and that both the running cost L and the terminal cost G are symmetric with respect to the configuration of agents. For player i, the system can be represented by a PDE that incorporates the distribution of the remaining $N-1$ players. Defining μ_i as the empirical measure that approximates the distribution of other agents:

$$\mu_i = \frac{1}{N-1} \sum_{j \neq i} \delta_{x_j},$$

we can write a deterministic equation as:

$$\begin{aligned} &\partial_t V(x,t,\mu_i) - H(x, \nabla_x V(x,t,\mu_i), \mu_i) \\ &- \int \partial_\mu V(x,t,\mu_i,\tilde{x}) \cdot \nabla_p H(\tilde{x}, \nabla_x V(\tilde{x},t,\mu_{\tilde{x}}), \mu_{\tilde{x}})\, d\mu_i(\tilde{x}) = 0. \end{aligned}$$

In this formulation:

- $V(x,t,\mu_i)$ represents the value function that provides the minimal cost starting from point x at time t under the distribution μ_i of other agents.

- The Hamiltonian $H(x, p, \mu)$ incorporates the interaction between the agent's own state x, momentum p, and the population distribution μ.
- The term $\partial_\mu V$ denotes the derivative of V with respect to the measure μ, capturing how the population distribution influences the optimal cost.

As N approaches infinity, this discrete empirical measure μ_i converges to a continuum distribution μ, leading to the mean field limit:

$$\partial_t V(x,t,\mu) - H(x, \nabla_x V(x,t,\mu), \mu) - \int \partial_\mu V(x,t,\mu,\tilde{x}) \cdot \nabla_p H(\tilde{x}, \nabla_x V(\tilde{x},t,\mu), \mu)\, d\mu(\tilde{x}) = 0.$$

This master equation describes the evolution of the value function in a continuum of agents, where each agent's optimal strategy depends on both its own state and the population distribution μ. This equation is central to deterministic mean field games, as it encapsulates the interactions within a large population without requiring detailed tracking of individual agents.

In this chapter we will show existence of unique classical solutions to the deterministic master equation under displacement monotonicity conditions and regularity assumptions (described formally in the next section).

2.2 Notation and Setup

Before proceeding to the master equation we give a brief refesher on the Wasserstein space and the Wasserstein calculus.

2.2.1 Elements of analysis and calculus on the Wasserstein space

Let $\mathscr{P}(\mathbb{R}^d)$ be the set of Borel probability measures supported in $\mathbb{R}^d$. For any $q \geq 1$ and any measure $\mu \in \mathscr{P}(\mathbb{R}^d)$, we set $M_q(\mu) := \left(\int_{\mathbb{R}^d} |x|^q d\mu(x)\right)^{\frac{1}{q}}$. Furthermore, let $\mathscr{P}_q(\mathbb{R}^d) := \{\mu \in$

$\mathscr{P}(\mathbb{R}^d): \ M_q(\mu) < \infty\}$. For any $\mu, \nu \in \mathscr{P}_q(\mathbb{R}^d)$, the W_q–Wasserstein distance is defined as

$$W_q(\mu, \nu) := \inf \Big\{ \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 d\gamma(x,y) : \ \gamma \in \Pi(\mu,\nu) \Big\}^{\frac{1}{q}},$$

where $\Pi(\mu,\nu) := \big\{ \gamma \in \mathscr{P}_2(\mathbb{R}^d \times \mathbb{R}^d) : \ (\pi_1)_\sharp \gamma = \mu, \ (\pi_2)_\sharp \gamma = \nu \big\}$, and $\pi_1, \pi_2 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ stand for the canonical projections, i.e. $\pi_1(x,y) = x$ and $\pi_2(x,y) = y$.

According to the terminology in [AGS08], the Wasserstein gradient of a function $U : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ at μ, is an element $\partial_\mu U(\mu, \cdot) \in \overline{\nabla C_c^\infty(\mathbb{R}^d)}^{L^2_\mu}$ (the closure of gradients of C_c^∞ functions in $L^2_\mu(\mathbb{R}^d;\mathbb{R}^d)$) and so, it is a priori defined μ–almost everywhere. The theory developed in [CP20a, GT19, Lio12b] shows that $\partial_\mu U(\mu, \cdot)$ can be characterized by the property

$$U(\mathcal{L}_{\xi+\eta}) - U(\mu) = \mathbb{E}\big[\langle \partial_\mu U(\mu, \xi), \eta \rangle\big] + o(\|\eta\|_2), \ \forall \ \xi, \eta, \ \textit{with } \mathcal{L}_\xi = \mu. \tag{2.1}$$

Let $\mathcal{C}^0(\mathscr{P}_2(\mathbb{R}^d))$ denote the space of W_2–continuous functions $U : \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$. For $k \in \{1,2\}$ we next define a subset of $\mathcal{C}^k(\mathscr{P}_2(\mathbb{R}^d))$, referred to as functions of full $\mathcal{C}^k$ regularity in [CD18a, Chapter 5]), as follows. By $\mathcal{C}^1(\mathscr{P}_2(\mathbb{R}^d))$, we mean the space of functions $U \in \mathcal{C}^0(\mathscr{P}_2(\mathbb{R}^d))$ such that $\partial_\mu U$ exists for all $\mu \in \mathscr{P}_2$ and it has a unique jointly continuous extension to $\mathscr{P}_2 \times \mathbb{R}^d$, which we continue to denote by

$$\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \ni (\tilde{x}, \mu) \mapsto \partial_\mu U(\mu, \tilde{x}) \in \mathbb{R}^d.$$

Similarly, $\mathcal{C}^2(\mathscr{P}_2(\mathbb{R}^d))$ stands for the space of functions $U \in \mathcal{C}^1(\mathscr{P}_2(\mathbb{R}^d))$ such that the global version of $\partial_\mu U$ is differentiable in the sense that all the following maps exist and have unique jointly continuous extensions

$$\begin{aligned} &\mathbb{R}^d \times \mathscr{P}_2 \ni (\tilde{x}, \mu) \mapsto \partial_{\tilde{x}\mu} U(\mu, \tilde{x}) \in \mathbb{R}^d \ \ \textit{and} \\ &\mathbb{R}^{2d} \times \mathscr{P}_2 \ni (\tilde{x}, \bar{x}, \mu) \mapsto \partial_{\mu\mu} U(\mu, \tilde{x}, \bar{x}) \in \mathbb{R}^{d\times d}. \end{aligned}$$

We define similarly the spaces $\mathcal{C}^1(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d))$ and $\mathcal{C}^2(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d))$. In particular $\mathcal{C}^2(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d))$ is the space of continuous functions $U : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ satisfying the following

(i) $\partial_x U, \partial_{xx} U$ exist and are jointly continuous on $\mathbb{R}^d \times \mathscr{P}_2$;

(ii) The following maps exist and have unique jointly continuous extensions

$$\mathbb{R}^{2d} \times \mathscr{P}_2 \ni (x, \tilde{x}, \mu) \mapsto \partial_\mu U(x, \mu, \tilde{x}) \in \mathbb{R}^d \quad \textit{and}$$
$$\mathbb{R}^{2d} \times \mathscr{P}_2 \ni (x, \tilde{x}, \mu) \mapsto \partial_{x\mu} U(x, \mu, \tilde{x}) \in \mathbb{R}^{d\times d};$$

(iii) Finally, the following maps exist and have unique jointly continuous extensions

$$\mathbb{R}^{2d} \times \mathscr{P}_2 \ni (x, \tilde{x}, \mu) \mapsto \partial_{\tilde{x}\mu} U(x, \mu, \tilde{x}) \in \mathbb{R}^{d\times d} \quad \text{and}$$
$$\mathbb{R}^{3d} \times \mathscr{P}_2 \ni (x, \tilde{x}, \bar{x}, \mu) \mapsto \partial_{\mu\mu} U(x, \mu, \tilde{x}, \bar{x}) \in \mathbb{R}^{d\times d}.$$

We underline that for notational conventions, we always denote the 'new spacial variables' appearing in Wasserstein derivatives with tilde symbols (for first order Wasserstein derivatives), with "bar" symbols (for second order Wasserstein derivatives) and so on, and we place them right after the corresponding measures variables. For example, when $U : \mathbb{R}^d \times \mathscr{P}_2 \times \mathbb{R}^d \to \mathbb{R}$ is typically evaluated as $U(x, \mu, p)$, we use the notations $\partial_\mu U(x, \mu, \tilde{x}, p)$, $\partial_{\tilde{x}} \partial_\mu U(x, \mu, \tilde{x}, p)$, $\partial_\mu \partial_\mu U(x, \mu, \tilde{x}, \bar{x}, p)$, and so on. This convention will be carried through to compositions with random variables too, for example $\partial_\mu U(x, \mu, \tilde{\xi}, p)$, when $\tilde{\xi}$ is an $\mathbb{R}^d$-valued random variable.

In this chapter we consider the master equation (1.1) with $\beta = \beta_0 = 0$. For the convenience of the reader we reproduce it here:

$$\begin{aligned} \partial_t V(t, x, \mu) + H(x, \mu, \partial_x V) - \mathcal{N} V(t, x, \mu) &= 0 \\ V(T, x, \mu) &= G(x, \mu) \end{aligned} \tag{2.2}$$

where

$$\mathcal{N} V(t, x, \mu) = -\int \partial_\mu V(t, x, \mu, \tilde{x}) \cdot \partial_p H(\tilde{x}, \mu, \partial_x V(t, \tilde{x}, \mu)) d\mu(\tilde{x})$$

We assume that we have a fixed atomless probability space $(\Omega, \mathbb{P})$. We use $L^2(\Omega)$ to denote the L^2 functions that map Ω into $\mathbb{R}^d$.

2.2.2 Displacement Monotonicity

Following [GMMZ22], we can recall the followings.

Definition 2.2.1. Let $\xi, \eta \in L^2(\Omega)$ and $U : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$. Assume that U is differentiable in the first variable and $\partial_x U \in \mathcal{C}^1(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d))$. We define

$$\begin{aligned}&(d_x d)_\xi U(\eta, \eta)\\ &:= \iint_{\Omega\times\Omega} \langle \partial_{x\mu} U\left(\xi(\omega), \mathcal{L}_\xi, \xi(\tilde{\omega})\right) \eta(\tilde{\omega}), \eta(\omega)\rangle + \langle \partial_{xx} U\left(\xi(\omega), \mathcal{L}_\xi\right) \eta(\omega), \eta(\omega)\rangle \, d\mathbb{P}(\omega) d\mathbb{P}(\tilde{\omega})\end{aligned}$$

We say that U is *displacement monotone* if $(d_x d)_\xi U(\eta(\omega), \eta(\omega)) \geq 0$ for all $\xi, \eta \in L^2(\Omega)$.

Remark 2.2.2. The condition can also be written directly in terms of measures. U is displacement monotone if and only if for every $\lambda \in \mathscr{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ we have

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} \langle \partial_{x\mu} U\left(x, (\pi_1)_\sharp \lambda, \tilde{x}\right) \tilde{y}, y\rangle + \langle \partial_{xx} U\left(x, (\pi_1)_\sharp\lambda\right) y, y\rangle \, d\lambda(x,y) d\lambda(\tilde{x}, \tilde{y}) \geq 0$$

where $(\pi_1)_\sharp \lambda$ stands for the left marginal of λ.

We recall the following Lemma from [GMMZ22, Lemma 2.6]. For the convenience of the reader we provide a detailed proof.

Lemma 2.2.3. *Suppose that U is displacement monotone and twice differentiable in space. If $\partial_{xx} U$ is continuous (where continuity in the measure variable is respect to W_2) then $\partial_{xx} U(x, \mu) \geq 0$, i.e. U is convex in space.*

Proof. Since absolutely continuous measures with positive and bounded densities are dense we may assume without loss of generality that $\mu = \rho dx$ with ρ positive and bounded (say $|\rho| \leq M$).

Fix some $x_0, z \in \mathbb{R}^d$ and let $\epsilon > 0$. Let ξ be distributed according to μ and let

$$\eta(\omega) = \begin{cases} 0, & \text{if } |\xi(\omega) - x_0| \geq \epsilon \\ \epsilon^{-d/2} z, & \text{else} \end{cases}$$

We now have

$$
\begin{aligned}
0 &\le (d_x d)_\xi U(\eta,\eta) \\
&= \iint_{\Omega\times\Omega} \langle \partial_{x\mu} U\left(\xi(\omega), \mathcal{L}_\xi, \xi(\tilde\omega)\right) \eta(\tilde\omega), \eta(\omega)\rangle + \langle \partial_{xx} U\left(\xi(\omega), \mathcal{L}_\xi\right) \eta(\omega), \eta(\omega)\rangle d\mathbb{P}(\omega) d\mathbb{P}(\tilde\omega) \\
&= \iint_{B(x_0,\epsilon)\times\Omega} \langle \partial_{x\mu} U\left(x, \mu, \xi(\tilde\omega)\right) \eta(\tilde\omega), \epsilon^{-d/2} z\rangle + \langle \partial_{xx} U\left(x, \mu\right) \epsilon^{-d/2} z, \epsilon^{-d/2} z\rangle \rho(x) dx d\mathbb{P}(\tilde\omega) \\
&= \iint_{B(x_0,\epsilon)\times B(x_0,\epsilon)} \langle \partial_{x\mu} U\left(x, \mu, \tilde x\right) \epsilon^{-d/2} z, \epsilon^{-d/2} z\rangle + \langle \partial_{xx} U\left(x, \mu\right) \epsilon^{-d/2} z, \epsilon^{-d/2} z\rangle \rho(x)\rho(\tilde x) dx d\tilde x \\
&= \epsilon^{-d} \iint_{B(x_0,\epsilon)\times B(x_0,\epsilon)} \langle \partial_{x\mu} U\left(x, \mu, \tilde x\right) z, z\rangle \rho(x)\rho(\tilde x) dx d\tilde x \\
&+ \epsilon^{-d} \int_{B(x_0,\epsilon)} \langle \partial_{xx} U\left(x, \mu\right) z, z\rangle \rho(x) dx
\end{aligned}
$$

Now we send $\epsilon \to 0$. We see that the first term goes to 0 (it is on the order of $\epsilon^{-d}\epsilon^{2d}$) where as the second term goes to $C\langle \partial_{xx} U\left(x_0, \mu\right) z, z\rangle$ where C is the volume of the unit ball in $\mathbb{R}^d$. Hence we have

$$
\langle \partial_{xx} U\left(x_0, \mu\right) z, z\rangle \ge 0
$$

as desired. □

Definition 2.2.4. Let $H : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$. Assume that H is differentiable in the x and p variables and $\partial_x H(\cdot,\cdot,p) \in \mathcal{C}^1(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d))$ for all $p \in \mathbb{R}^d$ and $\partial_p H(x,\cdot,\cdot) \in \mathcal{C}^1(\mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d)$ for all $x \in \mathbb{R}^d$. Let $\varphi \in C^1(\mathbb{R}^d;\mathbb{R}^d)$ be a bounded Lipschitz function, $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, and $\xi \in L^2(\Omega)$. We define

$$
\begin{aligned}
&(displ_\xi^\varphi H)(\eta,\eta) \\
&:= \iint_{\Omega\times\Omega} \Big[\Big\langle \partial_{x\mu} H(\xi(\omega), \mu, \xi(\tilde\omega), \varphi(\xi(\omega)))\eta(\tilde\omega) \\
&+ \partial_{xx} H(\xi(\omega), \mu, \varphi(\xi(\omega)))\eta(\omega),\ \eta(\omega) \Big\rangle \Big] d\mathbb{P}(\omega) d\mathbb{P}(\tilde\omega) + Q_\xi^\varphi H(\eta,\eta),
\end{aligned}
$$

where

$$Q_\xi^\varphi H(\eta,\eta)$$
$$:= \frac{1}{4}\int_\Omega \left[\left|\left(\partial_{pp}H(\xi(\omega),\mu,\varphi(\xi(\omega)))\right)^{-\frac{1}{2}}\int_\Omega \left[\partial_{p\mu}H(\xi(\omega),\mu,\xi(\tilde\omega),\varphi(\xi(\omega)))\eta(\tilde\omega)\right]d\mathbb{P}(\tilde\omega)\right|^2\right]d\mathbb{P}(\omega)$$

H is said to be *displacement monotone* if $(displ_\xi^\varphi H)(\eta,\eta) \le 0$ for all φ,μ,ξ.

2.3 Assumptions

We will always assume that G, H are displacement monotone. We also make the following regularity assumptions.

Assumption 1. We assume that

1. G is smooth.
2. $|\partial_x G|, |\partial_{xx} G|, |\partial_{\mu x} G|$ are uniformly bounded by L_0^G.

Assumption 2. We assume that

1. H is smooth.
2. All of the derivatives of H (not including H itself) are uniformly bounded in x, μ and locally in p.
3. $|\partial_x H(x,\mu,p)| \le C_1(1+|p|)$.
4. $\partial_{pp} H \ge c_0 I$ for some $c_0 > 0$.

We remark that 4 in the above tells us that H is convex in the p variable.

Definition 2.3.1. A constant C is said to be universal if it depends only on the above quantities (L_0^G and the bounds on H) and T.

We now translate these assumptions onto L.

Proposition 2.3.2. *Under the assumptions in 2 we have that*

1. *L is smooth.*

2. *$\partial_{vv} L \leq c_0^{-1} I$.*

3. *$|\partial_x L(x,\mu,v)| \leq C(1 + |\partial_v L(x,\mu,v)|)$.*

4. *$|\partial_v L(x,\mu,v)| \leq C(1+|v|)$.*

5. *$|\partial_x L(x,\mu,v)| \leq C(1+|v|)$.*

Proof. 1,2 are well known.

3 comes from the formula

$$\partial_x L(x,\mu,v) = -\partial_x H(x,\mu,\partial_v L(x,\mu,v)).$$

Now to prove 4 note that Assumption 2 on H implies that

$$|\partial_p H(x,\mu,0)| \leq C$$

and so

$$|\partial_p H(x,\mu,p)| \geq c_0 |p| - |\partial_p H(x,\mu,0)| \geq c_0 |p| - C$$

Hence

$$|v| = |\partial_p H(x,\mu,\partial_v L(x,\mu,v))| \geq c_0 |\partial_v L(x,\mu,v)| - C$$

as desired.

5 now follows from 3 and 4. □

Assumption 3. There is a superlinear function θ so that

$$L(x,\mu,v) \geq L(x,\mu,0) + \theta(|v|).$$

Lemma 2.3.3. *Let $C(R) = \sup_{x,\mu,p:|p|\le R} \|\partial_{pp} H(x,\mu,p)\|$.*

The above assumption holds if $\int_1^\infty C(R)^{-1}$ diverges.

Proof. We have

$$\begin{aligned}
L(x,\mu,v) &= L(x,\mu,0) + \int_0^1 \partial_v L(x,\mu,vs)\cdot v ds \\
&= L(x,\mu,0) + \int_0^1 \left(\partial_v L(x,\mu,0) + \int_0^s \partial_{vv} L(x,\mu,vr) v dr\right)\cdot v ds \\
&= L(x,\mu,0) + \partial_v L(x,\mu,0)\cdot v + \int_0^1\int_0^s \partial_{vv} L(x,\mu,vr) v\cdot v dr ds \\
&= L(x,\mu,0) + \partial_v L(x,\mu,0)\cdot v + \int_0^1\int_r^1 \partial_{vv} L(x,\mu,vr) v\cdot v ds dr \\
&= L(x,\mu,0) + \partial_v L(x,\mu,0)\cdot v + \int_0^1 (1-r)\partial_{vv} L(x,\mu,vr) v\cdot v dr \\
&\ge L(x,\mu,0) + \partial_v L(x,\mu,0)\cdot v + \frac{|v|}{2}\int_0^{\frac{|v|}{2}} \partial_{vv} L(x,\mu,r\hat{v})\hat{v}\cdot\hat{v} dr \\
&= L(x,\mu,0) + \partial_v L(x,\mu,0)\cdot v + \frac{|v|}{2}\int_0^{\frac{|v|}{2}} \partial_{pp} H(x,\mu,\partial_v L(x,\mu,r\hat{v}))^{-1}\hat{v}\cdot\hat{v} dr \\
&\ge L(x,\mu,0) + \partial_v L(x,\mu,0)\cdot v + \frac{|v|}{2}\int_0^{\frac{|v|}{2}} C^{-1}_{|\partial_v L(x,\mu,r\hat{v})|} dr \\
&\ge L(x,\mu,0) + \partial_v L(x,\mu,0)\cdot v + \frac{|v|}{2}\int_0^{\frac{|v|}{2}} C^{-1}_{Cr} dr \\
&= L(x,\mu,0) + \partial_v L(x,\mu,0)\cdot v + \frac{|v|}{2C}\int_0^{\frac{C|v|}{2}} C^{-1}_{r} dr
\end{aligned}$$

and so defining

$$\theta(|v|) = \frac{|v|}{2C}\int_0^{\frac{C|v|}{2}} C_r^{-1} dr - |\partial_v L(x,\mu,0)|\,|v|$$

we see that $L(x,\mu,v) \ge L(x,\mu,0) + \theta(|v|)$ and θ is superlinear. □

2.4 Propagation of Displacement Monotonicity

In this section we show that the conditions on H propagate the displacement monotonicity of G to V for all times. The proofs of this section are very similar to the proofs in [GMMZ22] however since the formal statements in that work require individual noise and the proofs are written in the notation of stochastic analysis we rewrite the proofs here in simplified notation.

Lemma 2.4.1. *Let* $Y, \tilde{Y} : M \to \mathbb{R}^d$ *where* $M = \mathbb{R}^d$ *or* $M = \Omega$ *depending on context. Then*

$$W^1(Y_{\#}\mu, \tilde{Y}_{\#}\mu) \leq \|Y - \tilde{Y}\|_\infty$$

Proof. We have

$$\begin{aligned} W^1(Y_{\#}\mu, \tilde{Y}_{\#}\mu) &= \sup_\phi \int \phi(x) d(Y_{\#}\mu - \tilde{Y}_{\#}\mu) \\ &= \sup_\phi \int \phi(Y(x)) - \phi(\tilde{Y}(x)) d\mu \\ &\leq \sup_\phi \int \left|\phi(Y(x)) - \phi(\tilde{Y}(x))\right| d\mu \\ &\leq \int \left|Y(x) - \tilde{Y}(x)\right| d\mu \\ &\leq \|Y - \tilde{Y}\|_\infty \end{aligned}$$

where the supremums are taken over ϕ that are 1-Lipschitz. □

Lemma 2.4.2. *Let* V *be a classical solution of the master equation* (2.2) *with* $V_x, V_{xx}, V_{x\mu}$ *uniformly bounded. Fix some* $\xi, \eta \in L^2(\Omega)$. *Then the following system of ODE's has a unique solution*

$$X_t(\omega) = \xi - \int_0^t H_p(X_s, \mu_s, \partial_x V(s, X_s, \mu_s))ds, \qquad \mu_t := X_t(\omega)\#\mathbb{P}$$
$$\delta X_t(\omega) = \eta - \int_0^t H_{px}(X_s)\delta X_s + \frac{1}{2}\int_\Omega [H_{p\mu}(X_t(\omega), X_t(\tilde{\omega}))\delta X_t(\tilde{\omega})]d\mathbb{P}(\tilde{\omega}) + H_{pp}(X_s)N_s ds$$

where

$$\begin{aligned} N_t := &\int_\Omega \partial_{x\mu} V(X_t(\omega), X_t(\tilde{\omega}))\delta X_t(\tilde{\omega})d\mathbb{P}(\tilde{\omega}) + \partial_{xx}V(X_t(\omega))\delta X_t(\omega) \\ &+ \frac{1}{2}H_{pp}(X_t(\omega))^{-1}\int_\Omega H_{p\mu}(X_t(\omega), X_t(\tilde{\omega}))\delta X_t(\tilde{\omega})d\mathbb{P}(\tilde{\omega}). \end{aligned}$$

Proof. Note that the first equation is decoupled from the second. Hence we first show that the first equation has a unique solution.

We prove this with the contraction mapping principle. Let B be the space of all $X : [0, T_1] \times \Omega \to \mathbb{R}^d$ so that $\|X - \xi\|_\infty < +\infty$. where T_1 will be chosen later. We equip B with the supremum metric given by $d(X, \tilde{X}) = \|X - \tilde{X}\|_\infty$. Define $A : B \to B$ by

$$A(X)_t = \xi + \int_0^t H_p(X_s, \mu_s, \partial_x V(t, X_s, \mu_s))ds.$$

Clearly X is a solution to the above system (for $t \in [0, T_1]$) if and only if X is a fixed point of A. We prove that A is a contraction. We have

$$\begin{aligned}\left|A(X)_t - A(\tilde{X})_t\right| &\leq \int_0^t \left|H_p(X_s, \mu_s, \partial_x V(t, X_s, \mu_s)) - H_p(\tilde{X}_s, \tilde{\mu}_s, \partial_x V(t, \tilde{X}_s, \tilde{\mu}_s))\right| ds \\ &\leq C \int_0^t \left|X_s - \tilde{X}_s\right| + W^1(\mu_s, \tilde{\mu}_s)ds\end{aligned}$$

where C depends on the bounds on the derivatives of V (specifically on $V_x, V_{xx}, V_{x\mu}$). By Lemma 2.4.1

$$C \int_0^t \left|X_s - \tilde{X}_s\right| + W^1(\mu_s, \tilde{\mu}_s)ds \leq C \int_0^t \left|X_s - \tilde{X}_s\right| + \|X_s - \tilde{X}_s\|_\infty ds \leq 2CT_1 \|X - \tilde{X}\|_\infty$$

so by taking $T_1 < \frac{1}{2C}$ we will get that A is a contraction. Since C depends only derivatives of V that are a priori uniformly bounded we may repeat this procedure for the whole time interval. Hence we obtain that the first equation has a unique solution which we denote X_t.

Now given X_t note that the second equation is just a linear ODE and so it will have a unique solution by another routine application of the contraction mapping principle. □

Remark 2.4.3. The second equation can be rewritten as

$$\delta X_t(\omega) = \eta - \int_0^t H_{px}(X_s)\delta X_s + \int_\Omega [H_{p\mu}(X_t(\omega), X_t(\tilde{\omega}))\delta X_t(\tilde{\omega})]d\mathbb{P}(\tilde{\omega}) + H_{pp}(X_s)\tilde{N}_s ds$$

where

$$\tilde{N}_t := \int_\Omega \partial_{x\mu} V(X_t(\omega), X_t(\tilde{\omega}))\delta X_t(\tilde{\omega})d\mathbb{P}(\tilde{\omega}) + \partial_{xx} V(X_t(\omega))\delta X_t(\omega).$$

Theorem 2.4.4. *Assume that G, H are displacement monotone and that they satisfy the regularity assumptions 1 and 2. Let V be a classical solution of the master equation* (2.2)*. Assume $\partial_x V \in C^{1,1}$ jointly in all variables and for each fixed t, we have $\partial_x V(t, \cdot) \in C^2$. Then $V(t, \cdot, \cdot)$ is displacement monotone for all $t \in [0, T]$.*

Proof. The proof of this Theorem follows the same argument of [GMMZ22] although we give the full details.

Without loss of generality, it suffices to prove the claim when $t = 0$.

Fix some $\xi, \eta \in L^2(\Omega)$ and let $X, \delta X$ be the unique solution to the ODE system from Lemma 2.4.2.

Define

$$I(t) := \iint_{\Omega\times\Omega} \langle \partial_{x\mu} V(X_t(\omega), X_t(\tilde{\omega}))\delta X_t(\tilde{\omega}), \delta X_t(\omega)\rangle d\mathbb{P}(\omega) d\mathbb{P}(\tilde{\omega})$$
$$\bar{I}(t) := \int_{\Omega} \langle \partial_{xx} V(X_t(\omega))\delta X_t(\omega), \delta X_t(\omega)\rangle d\mathbb{P}(\omega)$$

so that

$$I(t) + \bar{I}(t) = (d_x d)_{X_t(\omega)} V(\delta X_t(\omega), \delta X_t(\omega))$$

Since $V(T, \cdot) = G(\cdot)$ is displacement monotone we get that $I(T) + \bar{I}(T) \geq 0$. Note that when $t = 0$ we have $I(0) + \bar{I}(0) = (d_x d)_\xi V(\eta, \eta)$ and so it suffices to show that $I(0) + \bar{I}(0) \geq 0$ in order to prove that $V(0, \cdot)$ is displacement monotone. To do this we show $\dot{I}(t) + \dot{\bar{I}}(t) \leq 0$. We proceed by direct computation.

We have

$$\dot{I}(t) = I + II + III$$

where I comes from the terms where ∂_t hits V, II from when ∂_t hits $\delta X_t(\tilde{\omega})$, and III from when ∂_t hits $\delta X_t(\omega)$. We have

$$
\begin{aligned}
I := \iiint_{\Omega^3} & \langle \{\partial_{tx\mu} V(X_t(\omega), X_t(\tilde{\omega})) - H_p(X_t(\omega))^T \partial_{xx\mu} V(X_t(\omega), X_t(\tilde{\omega})) \\
& - H_p(X_t(\bar{\omega}))^T \partial_{\mu x\mu} V(X_t(\omega), X_t(\tilde{\omega}), X_t(\bar{\omega})) \\
& - H_p(X_t(\tilde{\omega}))^T \partial_{\tilde{x}x\mu} V(X_t(\omega), X_t(\tilde{\omega}))\} \delta X_t(\tilde{\omega}), \delta X_t(\omega) \rangle d\mathbb{P}(\omega) d\mathbb{P}(\tilde{\omega}) d\mathbb{P}(\bar{\omega}).
\end{aligned}
$$

Now rewrite $\tilde{\omega}$ in N as $\bar{\omega}$ and use the simplified expression in Remark 2.4.3 to obtain:

$$
\begin{aligned}
II &:= -\iiint_{\Omega^3} \langle \partial_{\mu x} V(X_t(\omega), X_t(\tilde{\omega})) \{[H_{px}(X_t(\omega)) \\
&\quad + H_{pp}(X_t(\omega)) \partial_{xx} V(X_t(\omega))] \delta X_t(\omega) + \mathbf{II_2}\}, \delta X_t(\tilde{\omega}) \rangle d\mathbb{P}(\omega) d\mathbb{P}(\tilde{\omega}) d\mathbb{P}(\bar{\omega}) \\
\mathbf{II_2} &:= [H_{p\mu}(X_t(\omega), X_t(\bar{\omega})) + H_{pp}(X_t(\omega)) \partial_{x\mu} V(X_t(\omega), X_t(\bar{\omega}))] \delta X_t(\bar{\omega}) \\
III &:= -\iiint_{\Omega^3} \langle \partial_{\mu x} V(X_t(\omega), X_t(\tilde{\omega})) \{[H_{px}(X_t(\tilde{\omega})) \\
&\quad + H_{pp}(X_t(\tilde{\omega})) \partial_{xx} V(X_t(\tilde{\omega}))] \delta X_t(\tilde{\omega}) + \mathbf{III_2}\}, \delta X_t(\omega) \rangle d\mathbb{P}(\omega) d\mathbb{P}(\tilde{\omega}) d\mathbb{P}(\bar{\omega}) \\
\mathbf{III_2} &:= [H_{p\mu}(X_t(\tilde{\omega}), X_t(\bar{\omega})) + H_{pp}(X_t(\tilde{\omega})) \partial_{x\mu} V(X_t(\tilde{\omega}), X_t(\bar{\omega}))] \delta X_t(\bar{\omega})
\end{aligned}
$$

Note that the III expression is the same as the II expression where all of the tildes have been toggled outside of the first $\partial_{\mu x} V$.

Next we apply $-\partial_{x\mu}$ to the master equation and rewrite $\tilde{x}$ in the nonlocal term, N, as $\bar{x}$ to obtain

$$
0 = -(\partial_{x\mu} \mathscr{L} V)(t, x, \mu, \tilde{x}) = J + JJ + JJJ. \tag{2.3}
$$

Here, we have set,

$$
\begin{aligned}
J :=& \partial_{tx\mu} V(x, \tilde{x}) - H_{x\mu}(x, \tilde{x}) - \partial_{xx} V(x) H_{p\mu}(x, \tilde{x}) \\
& -(H_{xp}(x) + \partial_{xx} V(x) H_{pp}(x)) \partial_{x\mu} V(x, \tilde{x}) - H_p(x)^T \partial_{xx\mu} V(x, \tilde{x}),
\end{aligned}
$$

$$
JJ := - H_p(\tilde{x})^T \partial_{\tilde{x}x\mu} V(x, \tilde{x}) - \partial_{x\mu} V(x, \tilde{x}) (H_{px}(\tilde{x}) + H_{pp}(\tilde{x}) \partial_{xx} V(\tilde{x}))
$$

and

$$
JJJ := \int_{\mathbb{R}^d} -H_p(\bar{x})^T \partial_{\mu x\mu} V(x, \tilde{x}, \bar{x}) - \partial_{x\mu} V(x, \bar{x}) [H_{p\mu}(\bar{x}, \tilde{x}) + H_{pp}(\bar{x}) \partial_{x\mu} V(\bar{x}, \tilde{x})] d\mu_t(\bar{x}).
$$

J is the term that arises when $-\partial_{x\mu}$ hits either V or H. Recall

$$\mathcal{N}V = \int_{\mathbb{R}^d} \partial_\mu V(t,x,\mu,\bar{x}) \cdot (\partial_p H)(\bar{x},\mu,\partial_x V(t,\bar{x},\mu))d\mu(\bar{x})$$

When $-\partial_{x\mu}$ hits this we get two terms. The first is JJ which is the gradient term where ∂_μ hits the integral itself. The second is JJJ which is the integral $\partial\mu$ term where the ∂_μ hits the integrand.

Now we evaluate (2.3) along $(X_t(\omega),\mu_t,X_t(\tilde{\omega}))$, multiply it by $\delta X_t(\tilde{\omega})$, and inner product with $\delta X_t(\omega)$. Note that since $\mu_t = X_t \# \mathbb{P}$ the integral over $\mathbb{R}^d$ against $\mu_t(\bar{x})$ in JJJ will become an integral over Ω and the $\bar{x}$ will become $X_t(\bar{\omega})$. Integrate J, JJ over $\omega,\tilde{\omega},\bar{\omega}$ and integrate JJJ over $\omega,\tilde{\omega}$. Hence we get

$$\begin{aligned} J \to \iiint_{\Omega^3} \langle [&\partial_{tx\mu} V(X_t(\omega),X_t(\tilde{\omega})) - H_{x\mu}(X_t(\omega),X_t(\tilde{\omega})) - \partial_{xx}V(X_t(\omega))H_{p\mu}(X_t(\omega),X_t(\tilde{\omega})) \\ &-(H_{xp}(X_t(\omega)) + \partial_{xx}V(X_t(\omega))H_{pp}(X_t(\omega)))\partial_{x\mu}V(X_t(\omega),X_t(\tilde{\omega})) \\ &- H_p(X_t(\omega))^T \partial_{xx\mu}V(X_t(\omega),X_t(\tilde{\omega}))]\delta X_t(\tilde{\omega}),\delta X_t(\omega)\rangle d\mathbb{P}(\omega)d\mathbb{P}(\tilde{\omega})d\mathbb{P}(\bar{\omega}), \end{aligned}$$

$$\begin{aligned} JJ \to -\iiint_{\Omega^3} \langle [&H_p(X_t(\tilde{\omega}))^T \partial_{\tilde{x}x\mu}V(X_t(\omega),X_t(\tilde{\omega})) \\ &- \partial_{x\mu}V(X_t(\omega),X_t(\tilde{\omega}))(H_{px}(X_t(\tilde{\omega})) \\ &+ H_{pp}(X_t(\tilde{\omega}))\partial_{xx}V(X_t(\tilde{\omega})))]\delta X_t(\tilde{\omega}),\delta X_t(\omega)\rangle d\mathbb{P}(\omega)d\mathbb{P}(\tilde{\omega})d\mathbb{P}(\bar{\omega}) \end{aligned}$$

and

$$\begin{aligned} JJJ \to \iiint_{\Omega^3} \langle [&-H_p(X_t(\bar{\omega}))^T \partial_{\mu x\mu}V(X_t(\omega),X_t(\tilde{\omega}),X_t(\bar{\omega})) \\ &- \partial_{x\mu}V(X_t(\omega),X_t(\bar{\omega}))\{H_{p\mu}(X_t(\bar{\omega}),X_t(\tilde{\omega})) \\ &+ H_{pp}(X_t(\bar{\omega}))\partial_{x\mu}V(X_t(\bar{\omega}),X_t(\tilde{\omega}))\}]\delta X_t(\tilde{\omega}),\delta X_t(\omega)\rangle d\mathbb{P}(\omega)d\mathbb{P}(\tilde{\omega})d\mathbb{P}(\bar{\omega}). \end{aligned}$$

Next we break up the JJJ term into two parts. The first is

$$\iiint_{\Omega^3} \langle -H_p(X_t(\bar{\omega}))^T \partial_{\mu x\mu}V(X_t(\omega),X_t(\tilde{\omega}),X_t(\bar{\omega}))\delta X_t(\tilde{\omega}),\delta X_t(\omega)\rangle d\mathbb{P}(\omega)d\mathbb{P}(\tilde{\omega})d\mathbb{P}(\bar{\omega})$$

which we leave as is. For the other part we make the change of variables that swaps $\tilde{\omega}$ and $\bar{\omega}$ giving:

$$\iiint_{\Omega^3} \langle [-\partial_{x\mu} V(X_t(\omega), X_t(\tilde{\omega}))[H_{p\mu}(X_t(\tilde{\omega}), X_t(\bar{\omega}))$$
$$+ H_{pp}(X_t(\tilde{\omega}))\partial_{x\mu} V(X_t(\tilde{\omega}), X_t(\bar{\omega}))] d\mu(X_t(\tilde{\omega}))] \delta X_t(\bar{\omega}), \delta X_t(\omega)\rangle d\mathbb{P}(\omega) d\mathbb{P}(\tilde{\omega}) d\mathbb{P}(\bar{\omega}).$$

We now subtract these expressions from our expression for $\dot{I}(t)$. In doing so we cancel many terms. Notice the second part of JJJ exactly cancels $\mathbf{III_2}$, I will be entirely eliminated from the first part of JJJ along with some terms from J and a term from JJ. Next the rest of JJ cancels III exactly since the $\mathbf{III_2}$ is already gone. At this point we are left with all of II along with some of J. We now transpose II (the part before $\mathbf{II_2}$) to swap the $\delta X_t(\omega)$ and $\delta X_t(\tilde{\omega})$ and use the symmetry properties (as in certain second derivative matrices are symmetric matrices) of the involved quantities to cancel a few more terms. Finally we are left with

$$\dot{I}(t) = \iiint_{\Omega^3} [-\langle \partial_{\mu x} V(X_t(\omega), X_t(\tilde{\omega})) \big[H_{p\mu}(X_t(\omega), X_t(\bar{\omega}))$$
$$+ H_{pp}(X_t(\omega))\partial_{x\mu} V(X_t(\omega), X_t(\bar{\omega}))\big] \delta X_t(\bar{\omega}), \delta X_t(\tilde{\omega})\rangle$$
$$+ \langle \big[H_{x\mu}(X_t(\omega), X_t(\tilde{\omega}))$$
$$+ \partial_{xx} V(X_t(\omega)) H_{p\mu}(X_t(\omega), X_t(\tilde{\omega}))\big] \delta X_t(\tilde{\omega}), \delta X_t(\omega)\rangle] d\mathbb{P}(\omega) d\mathbb{P}(\tilde{\omega}) d\mathbb{P}(\bar{\omega})$$

We now repeat the exact same procedure for $\bar{I}$. Differentiating in t we get

$$\dot{\bar{I}}(t) = \bar{I} + \overline{II}$$

where $\bar{I}$ comes from the terms where ∂_t hits V and $\overline{II}$ from when ∂_t hits either of the $\delta X_t(\omega)$. We have

$$\bar{I} := \iint_{\Omega^2} \langle \{ \partial_{txx} V(X_t(\omega)) - H_p(X_t(\omega))^T \partial_{xxx} V(X_t(\omega))$$
$$- H_p(X_t(\tilde{\omega}))^T \partial_{\mu xx} V(X_t(\omega), X_t(\tilde{\omega}))\} \delta X_t(\omega), \delta X_t(\omega)\rangle d\mathbb{P}(\omega) d\mathbb{P}(\tilde{\omega}).$$

Now using the simplified expression in Remark 2.4.3 we obtain:

$$\overline{II} := -2 \iint_{\Omega^2} \langle \partial_{xx} V(X_t(\omega)) \{ [H_{px}(X_t(\omega)) + H_{pp}(X_t(\omega)) \partial_{xx} V(X_t(\omega))] \delta X_t(\omega) \\ + [H_{p\mu}(X_t(\omega), X_t(\tilde{\omega})) + H_{pp}(X_t(\omega)) \partial_{x\mu} V(X_t(\omega), X_t(\tilde{\omega}))] \delta X_t(\tilde{\omega}) \}, \delta X_t(\omega) \rangle d\mathbb{P}(\omega) d\mathbb{P}(\tilde{\omega})$$

where the 2 in front is because there are two $\delta X_t(\omega)$ terms.

Next we apply $-\partial_{xx}$ to the master equation and rewrite $\tilde{x}$ in the nonlocal term, N, as $\bar{x}$ to obtain

$$0 = -(\partial_{xx} \mathscr{L} V)(t, x, \mu, \tilde{x}) = \bar{J} + \overline{JJ}. \tag{2.4}$$

Here, we have set,

$$\bar{J} := \partial_{txx} V(x) - H_{xx}(x) - 2\partial_{xx} V(x) H_{px}(x) - \partial_{xx} V(x) H_{px}(x) \partial_{xx} V(x) - H_p(x)^T \partial_{xxx} V(x)$$

and

$$\overline{JJ} := -\int_{\mathbb{R}^d} H_p(\tilde{x})^T \partial_{\mu xx} V(x, \tilde{x}) d\mu(\tilde{x})$$

Here $\bar{J}$ is the term that arises when $-\partial_{xx}$ hits either V or H and $\overline{JJ}$ is the term from when $-\partial_{xx}$ hits $\mathcal{N}$.

We evaluate the (2.4) along $(X_t(\omega), \mu_t)$ multiply it by $\delta X_t(\omega)$, and inner product with $\delta X_t(\omega)$. Just as in the case for I note that since $\mu_t = X_t \# \mathbb{P}$ the integral over $\mathbb{R}^d$ against $\mu(\tilde{x})$ in $\overline{JJ}$ will become an integral over Ω and the $\tilde{x}$ will become $X_t(\tilde{\omega})$. Integrate $\bar{J}$ over $\omega, \tilde{\omega}$ and integrate $\overline{JJ}$ over ω. We obtain

$$\bar{J} \to \iint_{\Omega^2} \langle [\partial_{txx} V(x) - H_{xx}(X_t(\omega)) - 2\partial_{xx} V(X_t(\omega)) H_{px}(X_t(\omega)) - \partial_{xx} V(X_t(\omega)) \\ H_{px}(X_t(\omega)) \partial_{xx} V(X_t(\omega)) - H_p(X_t(\omega))^T \partial_{xxx} V(X_t(\omega))] \delta X_t(\omega), \delta X_t(\omega) \rangle d\mathbb{P}(\omega) d\mathbb{P}(\tilde{\omega}),$$

$$\overline{JJ} \to -\iint_{\Omega^2} H_p(X_t(\tilde{\omega}))^T \partial_{\mu xx} V(X_t(\omega), X_t(\tilde{\omega})) d\mathbb{P}(\omega) d\mathbb{P}(\tilde{\omega})$$

We now subtract these expressions from $\bar{I}, \overline{II}$ to obtain

$$\dot{\bar{I}}(t) = \iint_{\Omega\times\Omega} [\langle H_{xx}(X_t(\omega)) - \partial_{xx}V(X_t(\omega))\{H_{pp}(X_t(\omega))\partial_{xx}V(X_t(\omega))\delta X_t(\omega)$$
$$- 2[H_{p\mu}(X_t(\omega), X_t(\tilde{\omega})) + H_{pp}(X_t(\omega))\partial_{x\mu}V(X_t(\omega), X_t(\tilde{\omega}))]\delta X_t(\tilde{\omega})\}, \delta X_t(\omega)\rangle] d\mathbb{P}(\omega)d\mathbb{P}(\tilde{\omega})$$
$$= \int_\Omega [-\langle H_{pp}(X_t(\omega))\partial_{xx}V(X_t(\omega))\delta X_t(\omega), \partial_{xx}V(X_t(\omega))\delta X_t(\omega)\rangle$$
$$- 2\langle H_{pp}(X_t(\omega))\partial_{xx}V(X_t(\omega))\delta X_t(\omega), \int_\Omega [\partial_{x\mu}V(X_t(\omega), X_t(\tilde{\omega}))\delta X_t(\tilde{\omega})] d\mathbb{P}(\tilde{\omega})\rangle$$
$$- 2\langle \partial_{xx}V(X_t(\omega))\delta X_t(\omega), \int_\Omega [H_{p\mu}(X_t(\omega), X_t(\tilde{\omega}))\delta X_t(\tilde{\omega})] d\mathbb{P}(\tilde{\omega})\rangle$$
$$+ \langle H_{xx}(X_t(\omega))\delta X_t(\omega), \delta X_t(\omega)\rangle] d\mathbb{P}(\omega).$$

Finally we combine our expressions for $\dot{I}$ and $\dot{\bar{I}}$ to obtain

$$\dot{I}(t) + \dot{\bar{I}}(t)$$
$$= \int_\Omega [-|H_{pp}^{\frac{1}{2}}(X_t(\omega))\{\int_\Omega [\partial_{x\mu}V(X_t(\omega), X_t(\tilde{\omega}))\delta X_t(\tilde{\omega})] d\mathbb{P}(\tilde{\omega}) + \partial_{xx}V(X_t(\omega))\delta X_t(\omega)\}|^2$$
$$- \langle \int_\Omega [H_{p\mu}(X_t(\omega), X_t(\tilde{\omega}))\delta X_t(\tilde{\omega})] d\mathbb{P}(\tilde{\omega}),$$
$$\int_\Omega [\partial_{x\mu}V(X_t(\omega), X_t(\tilde{\omega}))\delta X_t(\tilde{\omega})] d\mathbb{P}(\tilde{\omega}) + \partial_{xx}V(X_t(\omega))\delta X_t(\omega)\rangle$$
$$+ \langle \int_\Omega [H_{x\mu}(X_t(\omega), X_t(\tilde{\omega}))\delta X_t(\tilde{\omega})] d\mathbb{P}(\tilde{\omega}) + H_{xx}(X_t(\omega))\delta X_t(\omega), \delta X_t(\omega)\rangle] d\mathbb{P}(\omega)$$
$$= \int_\Omega [-|H_{pp}^{\frac{1}{2}}(X_t(\omega))\{\int_\Omega [\partial_{x\mu}V(X_t(\omega), X_t(\tilde{\omega}))\delta X_t(\tilde{\omega})] d\mathbb{P}(\tilde{\omega}) + \partial_{xx}V(X_t(\omega))\delta X_t(\omega)\}$$
$$+ \frac{1}{2}H_{pp}^{-\frac{1}{2}}(X_t(\omega)) \int_\Omega [H_{p\mu}(X_t(\omega), X_t(\tilde{\omega}))\delta X_t(\tilde{\omega})] d\mathbb{P}(\tilde{\omega})|^2$$
$$+ \langle \int_\Omega H_{x\mu}(X_t(\omega), X_t(\tilde{\omega}))\delta X_t(\tilde{\omega}) d\mathbb{P}(\tilde{\omega}), \delta X_t(\omega)\rangle + \langle H_{xx}(X_t(\omega))\delta X_t(\omega), \delta X_t(\omega)\rangle$$
$$+ \frac{1}{4}|H_{pp}^{-\frac{1}{2}}(X_t(\omega)) \int_\Omega [H_{p\mu}(X_t(\omega), X_t(\tilde{\omega}))\delta X_t(\tilde{\omega})] d\mathbb{P}(\tilde{\omega})|^2] d\mathbb{P}(\omega)$$
$$= -\int_\Omega [|H_{pp}^{\frac{1}{2}}(X_t(\omega))N_t|^2] d\mathbb{P}(\omega) + (displ^{\varphi}_{X_t(\omega)}H)(\delta X_t(\tilde{\omega}), \delta X_t(\tilde{\omega})), \tag{2.5}$$

where $\varphi(x) := \partial_x V(t, x, \mu_t(\omega^0))$ and the last line is from the definition of $displ^{\varphi}_{X_t(\omega)}$. Applying the fact that H is displacement monotone we obtain that

$$\dot{I}(t) + \dot{\bar{I}}(t) \leq 0$$

as desired. ∎

2.5 A-Priori Bounds in Space

This section represents the main fundamental departure from [GMMZ22]. In that work the a-priori bounds in space were a simple consequence of the parabolic PDE theory because of noise. Here we have a deterministic system and so we must work harder to obtain these bounds.

Lemma 2.5.1. *Let V be a classical solution to the master equation. Fix some $\mu \in \mathcal{P}_2$. Then there exists a path $m(t) : [0,T] \to \mathcal{P}_2$ with $m(0) = \mu$ so that the following holds. Let $U(t,x) = V(t,x,m(t))$. Then*

$$U(t,x) = \inf_{\gamma:\gamma(t)=x} \int_t^T L(\gamma(s), m(s), \dot{\gamma}(s))ds + G(\gamma(T)).$$

Remark 2.5.2. The above formula for U is called the representation formula. What this says is that once we fix an initial distribution μ then V becomes the value function for some optimal control problem with Lagrangian $\tilde{L}(t,x,v) = L(x,v,m(t))$.

Proof of Lemma 2.5.1. We have that U satisfies the Hamilton Jacobi equation

$$\partial_t U(t,x) + H(x, m(t), \partial_x U) = 0$$
$$U(T,x) = G(x)$$

By [CL86, Theorem 1] this Hamilton Jacobi equation has a unique solution (note that the assumptions here are just regularity assumptions on H which are implied by our assumptions). By [CS04, Theorem 6.4.5] we have that the claimed representation formula is a solution to this equation (note that the assumptions here are regularity assumptions along with convexity and superlinear growth in v for L which are implied by our assumptions). Hence U is equal to the claimed representation formula. ∎

Lemma 2.5.3. *Fix some $\tilde{t}, x$ and let γ be the optimizer in the representation formula for U. Then there exists a universal constant C so that $|\dot{\gamma}| \leq C$.*

Proof. We first would like to show that $\int_t^T |\dot{\gamma}| \, ds$ is bounded by a universal constant. To see this note that by Assumption 3 we have

$$\begin{aligned}
\int_t^T \theta(|\dot{\gamma}(s)|) + L(\gamma(s), m(s), 0) ds &\leq \int_t^T L(\gamma(s), m(s), \dot{\gamma}(s)) ds \\
&= \int_t^T L(\gamma(s), m(s), \dot{\gamma}(s)) ds + G(\gamma(T)) - G(\gamma(T)) \\
&\leq \int_t^T L(\gamma(t), m(s), 0) ds + G(\gamma(t)) - G(\gamma(T)) \\
&\leq \int_t^T L(\gamma(t), m(s), 0) ds + C |\gamma(T) - \gamma(t)|
\end{aligned}$$

where we have used that G is Lipschitz with universal Lipschitz constant in the last line (this is from Assumption 1). Hence

$$\begin{aligned}
\int_t^T \theta(|\dot{\gamma}(s)|) ds &\leq \int_t^T L(\gamma(t), m(s), 0) - L(\gamma(s), m(s), 0) ds + C |\gamma(T) - \gamma(t)| \\
&\leq C \int_t^T |\gamma(s) - \gamma(t)| \, ds + C |\gamma(T) - \gamma(t)| \\
&\leq C \int_t^T \int_t^s |\dot{\gamma}(r)| \, dr ds + C |\gamma(T) - \gamma(t)| \\
&\leq C \int_t^T |\dot{\gamma}(s)| \, ds
\end{aligned}$$

where the second inequality is from (5) in Proposition 2.3.2. Now there is some universal constant $\tilde{C}$ so that $\theta(v) \geq (C+1)|v| - \tilde{C}$. The claim now follows.

We proceed to the proof of the lemma. By Assumption 3 we have

$$\theta(|\dot{\gamma}(s)|) \leq L(\gamma(s), m(s), \dot{\gamma}(s)) - L(\gamma(s), m(s), 0) \leq L_v(\gamma(s), m(s), \dot{\gamma}(s)) \cdot \dot{\gamma}(s)$$

since L is convex. Continuing we get

$$\begin{aligned}\theta(|\dot\gamma(s)|) &\le L_v(\gamma(s), m(s), \dot\gamma(s)) \cdot \dot\gamma(s)\\ &= \left(L_v(\gamma(T), m(T), \dot\gamma(T)) - \int_s^T \frac{d}{dt} L_v(\gamma(t), m(t), \dot\gamma(t))dt\right) \cdot \dot\gamma(s)\\ &= \left(L_v(\gamma(T), m(T), \dot\gamma(T)) - \int_s^T L_x(\gamma(t), m(t), \dot\gamma(t))dt\right) \cdot \dot\gamma(s)\end{aligned}$$

where we have used the Euler-Lagrange equation for the last line. Now by Hamilton's equations we have

$$\begin{aligned}&\left(L_v(\gamma(T), m(T), \dot\gamma(T)) - \int_s^T L_x(\gamma(t), m(t), \dot\gamma(t))dt\right) \cdot \dot\gamma(s)\\ &= \left(L_v(\gamma(T), m(T), H_p(\gamma(T), m(T), \partial_x G(\gamma(T)))) - \int_s^T L_x(\gamma(t), m(t), \dot\gamma(t))dt\right) \cdot \dot\gamma(s)\\ &= \left(\partial_x G(\gamma(T)) - \int_s^T L_x(\gamma(t), m(t), \dot\gamma(t))dt\right) \cdot \dot\gamma(s)\end{aligned}$$

where the last line follows because H, L are Legendre transforms of each other. Finally applying Assumption 1 and (5) in Proposition 2.3.2 we get

$$\begin{aligned}\left(\partial_x G(\gamma(T)) - \int_s^T L_x(\gamma(t), m(t), \dot\gamma(t))dt\right) \cdot \dot\gamma(s) &\le \left(C + C\int_s^T 1 + |\dot\gamma(t)|\, dt\right) |\dot\gamma(s)|\\ &\le C + C\,|\dot\gamma(s)|\end{aligned}$$

All together we obtain $\theta(|\dot\gamma(s)|) \le C + C\,|\dot\gamma(s)|$.

Now there is some further universal constant $\tilde{C}$ so that $\theta(v) \ge (1 + C)v - \tilde{C}$. Using this we obtain $|\dot\gamma(s)| \le C + \tilde{C}$ as desired. □

Theorem 2.5.4. *Suppose that V is a classical solution to the master equation. Then $\partial_x V$ is bounded by a universal constant. Furthermore V is semi-concave with universal semi-concavity constant.*

Proof. We want to bound $\partial_x V(t, x, \mu)$. Without loss of generality we bound $\partial_x V(0, x, \mu)$. Fix some μ and consider the corresponding U from Lemma 2.5.1. We have $\partial_x V(0, x, \mu) =$

$\partial_x U(0,x)$. Let γ be the optimal path in the representation formula for $U(0,x)$. Let $z \in \mathbb{R}^n$ be arbitrary. We set $\gamma_z(s) = \gamma(s) + z$. We have

$$\begin{aligned}
&U(0,x+z) - U(0,x) \\
&= U(0,x+z) - \left(\int_0^T L(\gamma(s), m(s), \dot{\gamma}(s))ds + G(\gamma(T))\right) \\
&\leq \int_0^T L(\gamma_z(s), m(s), \dot{\gamma}_z(s))ds + G(\gamma_z(T)) - \left(\int_0^T L(\gamma(s), m(s), \dot{\gamma}(s))ds + G(\gamma(T))\right) \\
&= \int_0^T L(\gamma_z(s), m(s), \dot{\gamma}(s)) - L(\gamma(s), m(s), \dot{\gamma}(s))ds + G(\gamma(T)+z) - G(\gamma(T)) \\
&\leq \int_0^T |z| \sup_{\tilde{x},\tilde{\mu},v \in \mathbb{R}^n \times \mathcal{P}_2 \times B(C)} L_x(\tilde{x}, \tilde{\mu}, v)ds + G(\gamma(T)+z) - G(\gamma(T)) \\
&\leq C\,|z|
\end{aligned}$$

which shows that U is Lipschitz in the space variable.

Next we do the semi-concavity. The objective is to show that there is some universal $C > 0$ so that

$$V(t, x+z, \mu) - 2V(t,x,\mu) + V(t, x-z, \mu) \leq C\,|z|^2$$

Again without loss of generality we assume that $t = 0$ and we fix an initial μ and so the objective becomes to show

$$U(0, x+z) - 2U(0,x) + U(0, x-z) \leq C\,|z|^2$$

Let γ be the optimal path in the representation formula for $U(0,x)$. We have

$$
\begin{aligned}
&U(0,x+z) - 2U(0,x) + U(0,x-z) \\
&= U(0,x+z) + U(0,x-z) - 2\left(\int_0^T L(\gamma(s), m(s), \dot{\gamma}(s))ds + G(\gamma(T))\right) \\
&\leq \left(\int_0^T L(\gamma_z(s), m(s), \dot{\gamma}_z(s))ds + G(\gamma_z(T))\right) \\
&+ \left(\int_0^T L(\gamma_{-z}(s), m(s), \dot{\gamma}_{-z}(s))ds + G(\gamma_{-z}(T))\right) \\
&- 2\left(\int_0^T L(\gamma(s), m(s), \dot{\gamma}(s))ds + G(\gamma(T))\right) \\
&= \int_0^T L(\gamma(s)+z, m(s), \dot{\gamma}(s)) - 2L(\gamma(s), m(s), \dot{\gamma}(s)) + L(\gamma(s)-z, m(s), \dot{\gamma}(s))ds \\
&+ G(\gamma(T)+z) - 2G(\gamma(T)) + G(\gamma(T)-z) \\
&\leq \int_0^T \|\partial_{xx} L\|_{L^\infty(B_C)} |z|^2 ds + \|\partial_{xx} G\|_\infty |z|^2 \\
&\leq C|z|^2
\end{aligned}
$$

as desired. □

Corollary 2.5.5. *Suppose that V is a classical solution to the master equation. Then $\partial_{xx} V$ is bounded by a universal constant.*

Proof. From Theorem 2.5.4 we have that $\partial_{xx} V \leq CI$. But because V is displacement convex we have that $\partial_{xx} V \geq 0$ by Lemma 2.2.3. Hence the result follows. □

2.6 W_2 A-Priori Bounds in Measure

In this section we show that the conditions of displacement monotonicity gives bounds on $\partial_x V$ in the measure variable. The proofs of this section are very similar to the proofs in [GMMZ22] however since the formal statements in that work require individual noise and the proofs are written in the notation of stochastic analysis we rewrite the proofs here in simplified notation.

Theorem 2.6.1. *Suppose that V is a smooth solution to the master equation. Then $\partial_x V$ is Lipschitz in the measure variable with respect to W_2. In particular the Lipschitz constant is bounded by a universal constant.*

Proof. The proof is essentially the same as [GMMZ22, Theorem 5.1] however we give all the details.

Fix some $\xi, \eta \in L^2(\Omega)$. Let $X, \delta X$ be the solution to the system in Lemma 2.4.2. By integrating equation (2.5) in the proof of Theorem 2.4.4 over t, we have (keeping the same notation as in that proof)

$$
\begin{aligned}
&\int_0^t \int_\Omega \left| H_{pp}(X_s(\omega))^{\frac{1}{2}} N_s \right|^2 d\mathbb{P}(\omega) ds \\
&= [I(0) + \bar{I}(0)] - [I(t) + \bar{I}(t)] + \int_0^t (displ^{\varphi}_{X_s(\omega)} H)(\delta X_s(\tilde{\omega}), \delta X_s(\tilde{\omega})) ds \\
&= [I(0) + \bar{I}(0)] - [I(t) + \bar{I}(t)] + \int_0^t (displ^{\varphi}_{X_s(\omega)} H)(\delta X_s(\tilde{\omega}), \delta X_s(\tilde{\omega})) ds \\
&= [I(0) + \bar{I}(0)] - [I(t) + \bar{I}(t)] + \int_0^t \int_{\Omega^2} [\langle H_{x\mu}(X_s(\omega), X_s(\tilde{\omega})) \delta X_s(\tilde{\omega}), \delta X_s(\omega) \rangle \\
&+ \langle H_{xx}(X_s(\omega)) \delta X(\omega), \delta X(\omega) \rangle + \frac{1}{4} \left| H_{pp}^{-\frac{1}{2}}(X_s(\omega)) H_{p\mu}(X_s(\omega), X_s(\tilde{\omega})) \delta X_s(\tilde{\omega}) \right|^2] d\mathbb{P}(\omega) d\mathbb{P}(\tilde{\omega}) ds \\
&\leq I(0) - [I(t) + \bar{I}(t)] + C \int_\Omega |\eta|^2 \, d\mathbb{P}(\omega) + C \int_0^t \int_\Omega |\delta X_s(\omega)|^2 d\mathbb{P}(\omega) ds
\end{aligned}
$$

where we have used that by Theorem 2.5.4 and Corollary 2.5.5 $\|\partial_x V\|_\infty, \|\partial_{xx} V\|_\infty$ are bounded by a universal constant. In particular all the derivatives of H that appear are uniformly bounded by a universal constant (since we assumed all derivatives of H are locally bounded in p). Since V is displacement monotone we have that $I(t) + \bar{I}(t) \geq 0$ and so

$$
\begin{aligned}
c_0 \int_0^t \int_\Omega |N_s|^2 \, d\mathbb{P}(\omega) ds \leq \int_0^t \int_\Omega \left| H_{pp}(X_s(\omega))^{\frac{1}{2}} N_s \right|^2 d\mathbb{P}(\omega) ds \\
\leq I(0) + C \int_\Omega |\eta|^2 \, d\mathbb{P}(\omega) + C \int_0^t \int_\Omega |\delta X_s(\omega)|^2 d\mathbb{P}(\omega) ds
\end{aligned}
$$

Recall that the defining ODE for δX is

$$
\delta X_t = \eta - \int_0^t H_{px}(X_s) \delta X_s + \frac{1}{2} \int_\Omega [H_{p\mu}(X_s(\omega), X_s(\tilde{\omega})) \delta X_s(\tilde{\omega})] d\mathbb{P}(\tilde{\omega}) + H_{pp}(X_s) N_s ds
$$

and so

$$|\delta X_t|^2 \leq 2\,|\eta|^2 + C \int_0^t |\delta X_s|^2 + \int_\Omega |\delta X_s(\tilde{\omega})|^2 \, d\mathbb{P}(\tilde{\omega}) + |N_s|^2 \, ds.$$

Now integrating over ω we get

$$\begin{aligned}
\int_\Omega |\delta X_t|^2 \, d\mathbb{P}(\omega) &\leq 2 \int_\Omega |\eta|^2 \, d\mathbb{P}(\omega) + C \int_0^t \int_\Omega |\delta X_s|^2 \, d\mathbb{P}(\omega) ds + C \int_0^t \int_\Omega |N_s|^2 \, d\mathbb{P}(\omega) ds \\
&\leq C \int_\Omega |\eta|^2 \, d\mathbb{P}(\omega) + C \int_0^t \int_\Omega |\delta X_s|^2 \, d\mathbb{P}(\omega) ds + C\,|I(0)|
\end{aligned}$$

and so by Gronwall's inequality we have

$$\sup_{t\in[0,T]} \int_\Omega |\delta X_t|^2 \, d\mathbb{P}(\omega) \leq C \int_\Omega |\eta|^2 \, d\mathbb{P}(\omega) + C\,|I(0)|$$

We define

$$\Upsilon_t(\omega) = \int_\Omega \partial_{x\mu} V(t, X_t(\omega), \mu_t, X_t(\tilde{\omega}))\delta X(\tilde{\omega})_t d\mathbb{P}(\tilde{\omega})$$

Note that Υ_t is like $I(t)$ from the proof of Theorem 2.4.4 except that we don't inner product with δX. In particular

$$\begin{aligned}
|I(0)| &\leq \int_\Omega \left| \int_\Omega \langle \partial_{x\mu} V(X_0(\omega), X_0(\tilde{\omega}))\delta X_0(\tilde{\omega}), \eta(\omega)\rangle d\mathbb{P}(\tilde{\omega}) \right| d\mathbb{P}(\omega) \\
&= \int_\Omega \left| \int_\Omega \langle \partial_{x\mu} V(X_0(\omega), X_0(\tilde{\omega}))\delta X_0(\tilde{\omega}) d\mathbb{P}(\tilde{\omega}), \eta(\omega)\rangle \right| d\mathbb{P}(\omega) \\
&\leq \int_\Omega \left| \int_\Omega \partial_{x\mu} V(X_0(\omega), X_0(\tilde{\omega}))\delta X_0(\tilde{\omega}) d\mathbb{P}(\tilde{\omega}) \right| |\eta(\omega)| \, d\mathbb{P}(\omega) \\
&= \int_\Omega |\Upsilon_0(\omega)| \, |\eta(\omega)| \, d\mathbb{P}(\omega)
\end{aligned}$$

so that

$$\sup_{t\in[0,T]} \int_\Omega |\delta X_t|^2 \, d\mathbb{P}(\omega) \leq C \int_\Omega |\eta|^2 + |\eta|\,|\Upsilon_0| \, d\mathbb{P}(\omega)$$

Now in the exact same manner as in the proof of Theorem 2.4.4 we compute

$$\dot{\Upsilon}_t = K_3(t) - K_4(t)$$

where

$$
\begin{aligned}
K_3(t) &:= \iint_{\Omega^2} \{\partial_{tx\mu} V(X_t(\omega), X_t(\tilde{\omega})) - H_p(X_t(\omega))^T \partial_{xx\mu} V(X_t(\omega), X_t(\tilde{\omega})) \\
&\quad - H_p(X_t(\bar{\omega}))^T \partial_{\mu x\mu} V(X_t(\omega), X_t(\tilde{\omega}), X_t(\bar{\omega})) \\
&\quad - H_p(X_t(\tilde{\omega}))^T \partial_{\tilde{x}x\mu} V(X_t(\omega), X_t(\tilde{\omega}))\} \delta X_t(\tilde{\omega}) d\mathbb{P}(\tilde{\omega}) d\mathbb{P}(\bar{\omega}) \\
K_4(t) &:= -\iint_{\Omega^2} \langle \partial_{\mu x} V(X_t(\omega), X_t(\tilde{\omega})) \{[H_{px}(X_t(\tilde{\omega})) \\
&\quad + H_{pp}(X_t(\tilde{\omega})) \partial_{xx} V(X_t(\tilde{\omega}))] \delta X_t(\tilde{\omega}) + \mathbf{KK_2}\} d\mathbb{P}(\tilde{\omega}) d\mathbb{P}(\bar{\omega}) \\
\mathbf{KK_2} &:= [H_{p\mu}(X_t(\tilde{\omega}), X_t(\bar{\omega})) + H_{pp}(X_t(\tilde{\omega})) \partial_{x\mu} V(X_t(\tilde{\omega}), X_t(\bar{\omega}))] \delta X_t(\bar{\omega}).
\end{aligned}
$$

Note that these correspond to terms I, III in the expression for $\dot{I}(t)$. We may now use the exact same procedure of applying $-\partial_{x\mu}$ to the master equation, getting J, JJ, JJJ and subtracting. After this we get

$$
\begin{aligned}
K_3(t) - K_4(t) = \int_{\Omega} &\{H_{x\mu}(X_t(\omega), X_t(\tilde{\omega})) + \partial_{xx} V(X_t(\omega)) H_{p\mu}(X_t(\omega), X_t(\tilde{\omega})) \\
&+ [H_{xp}(X_t(\omega)) + \partial_{xx} V(X_t(\omega)) H_{pp}(X_t(\omega))] \partial_{x\mu} V(X_t(\omega), X_t(\tilde{\omega}))\} \delta X_t(\tilde{\omega}) d\mathbb{P}(\tilde{\omega})
\end{aligned}
$$

Let

$$
\begin{aligned}
K_5(t) &:= H_{xp}(X_t(\omega)) + \partial_{xx} V(X_t(\omega)) H_{pp}(X_t(\omega)) \\
K_6(t) &:= \int_{\Omega} \{H_{x\mu}(X_t(\omega), X_t(\tilde{\omega})) + \partial_{xx} V(X_t(\omega)) H_{p\mu}(X_t(\omega), X_t(\tilde{\omega}))\} \delta X_t(\tilde{\omega}) d\mathbb{P}(\tilde{\omega})
\end{aligned}
$$

So that $K_3(t) - K_4(t) = K_5(t)\Upsilon_t + K_6(t)$. Now we have $|K_5(t)| \le C$ and

$$
|K_6(t)| \le C \int_{\Omega} |\delta X_t(\tilde{\omega})| d\mathbb{P}(\tilde{\omega})
$$

for some universal constant C. Finally

$$
\begin{aligned}
\Upsilon_T &= \int_{\Omega} \partial_{x\mu} V(T, X_T(\omega), \mu_T, X_T(\tilde{\omega})) \delta X(\tilde{\omega})_T d\mathbb{P}(\tilde{\omega}) \\
&= \int_{\Omega} \partial_{x\mu} G(X_T(\omega), \mu_T, X_T(\tilde{\omega})) \delta X(\tilde{\omega})_T d\mathbb{P}(\tilde{\omega})
\end{aligned}
$$

and so

$$|\Upsilon_T| \leq C \int_\Omega |\delta X_T(\tilde{\omega})| d\mathbb{P}(\tilde{\omega})$$

Recall

$$\dot{\Upsilon}_t = K_5(t)\Upsilon_t + K_6(t)$$

so

$$\begin{aligned}|\Upsilon_t|^2 &\leq 2\,|\Upsilon_T|^2 + C\int_t^T [|\Upsilon_s|^2 + \int_\Omega |\delta X_t(\tilde{\omega})|^2 d\mathbb{P}(\tilde{\omega})]ds \\ &\leq C\int_\Omega |\delta X_T(\tilde{\omega})|^2 d\mathbb{P}(\tilde{\omega}) + C\int_t^T [|\Upsilon_s|^2 + \int_\Omega |\delta X_t(\tilde{\omega})|^2 d\mathbb{P}(\tilde{\omega})]ds\end{aligned}$$

Now Gronwall's inequality tells us that

$$|\Upsilon_t|^2 \leq C\int_\Omega |\delta X_T(\tilde{\omega})|^2 + C\int_0^T \int_\Omega |\delta X_t(\tilde{\omega})|^2 d\mathbb{P}(\tilde{\omega})ds$$

and so

$$|\Upsilon_0|^2 \leq C \sup_{t\in[0,T]} \int_\Omega |\delta X_t|^2 \, d\mathbb{P}(\omega)$$

Let $\epsilon > 0$ be arbitrary. Recall that we have shown

$$\sup_{t\in[0,T]} \int_\Omega |\delta X_t|^2 \, d\mathbb{P}(\omega) \leq C\int_\Omega |\eta|^2 + |\eta|\,|\Upsilon_0|\, d\mathbb{P}(\omega)$$

and so

$$\begin{aligned}\sup_{t\in[0,T]} \int_\Omega |\delta X_t|^2 \, d\mathbb{P}(\omega) &\leq \int_\Omega \frac{C}{\epsilon}\,|\eta|^2 + C\epsilon\,|\Upsilon_0|^2 \, d\mathbb{P}(\omega) \\ &\leq \int_\Omega \frac{C}{\epsilon}\,|\eta|^2 \, d\mathbb{P}(\omega) + C\epsilon \sup_{t\in[0,T]} \int_\Omega |\delta X_t|^2 \, d\mathbb{P}(\omega)\end{aligned}$$

where C is independent of ϵ. By taking $\epsilon = \frac{1}{2C}$ we obtain

$$\sup_{t\in[0,T]} \int_\Omega |\delta X_t|^2 \, d\mathbb{P}(\omega) \leq C\int_\Omega |\eta|^2 \, d\mathbb{P}(\omega)$$

and so

$$|\Upsilon_0(\omega)|^2 \leq C \int_\Omega |\eta|^2 \, d\mathbb{P}(\omega)$$

Recalling the definition of Υ_0 we have

$$\left| \int_\Omega \partial_{x\mu} V(0, \xi(\omega), \mu, \xi(\tilde{\omega}))\eta(\tilde{\omega}) d\mathbb{P}(\tilde{\omega}) \right|^2 \leq C \int_\Omega |\eta|^2 \, d\mathbb{P}$$

for every $\omega \in \Omega$. In particular for μ almost every x we get

$$\left| \int_\Omega \partial_{x\mu} V(0, x, \mu, \xi(\tilde{\omega}))\eta(\tilde{\omega}) d\mathbb{P}(\tilde{\omega}) \right|^2 \leq C \int_\Omega |\eta|^2 \, d\mathbb{P}$$

and so

$$\begin{aligned} |\partial_x V(0, x, \mathcal{L}_{\xi+\eta}) - \partial_x V(0, x, \mathcal{L}_\xi)| &= \left| \int_0^1 \int_\Omega \partial_{x\mu} V(0, x, \mathcal{L}_{\xi+\theta\eta}, \xi + \theta\eta)\eta d\mathbb{P} d\theta \right| \\ &\leq C \left(\int_\Omega |\eta|^2 \, d\mathbb{P} \right)^{\frac{1}{2}}. \end{aligned}$$

Now if ξ, η are chosen so $W_2^2(\mathcal{L}_{\xi+\eta}, \mathcal{L}_\xi) = \int_\Omega |\eta|^2 \, d\mathbb{P}$ then we will obtain the desired Lipschitz regularity of $\partial_x V$ in measure. □

2.7 Short Time Existence and Regularity

In this section we derive short time existence results for the master equation. Our methods are inspired by the short time results of [GM22a] however the lack of separability for our Hamiltonian introduces some new difficulties.

2.7.1 Hamiltonian ODE's

Proposition 2.7.1. *Suppose some G, μ are fixed. Then there is a $\delta > 0$ so that the system*

$$\begin{aligned}
\dot{Y}_t &= \partial_p H(Y_t, Y_t\#\mu, Z_t) \\
\dot{Z}_t &= -\partial_x H(Y_t, Y_t\#\mu, Z_t) \\
Y_0 &= \mathrm{Id} \\
Z_T &= \partial_x G(Y_T, Y_T\#\mu)
\end{aligned}$$

has a unique solution whenever $T < \delta$. Furthermore δ depends only on $\|G_x\|$, $\|G_{xx}\|$, the W_2 Lipschitz constant of G_x, and universal constants. Also $\|Z\|_\infty \leq 2\|G_x\|_\infty + 4C_G$ where $C_G = \|G_{xx}\| + C_{G\mu}$ and $C_{G\mu}$ is the W_2 Lipschitz constant of G_x with respect to the measure variable.

Here $Y_t, Z_t : \mathbb{R}^n \to \mathbb{R}^n$.

Proof. We first prove this under the additional assumption that $C_G \leq \frac{1}{4}$.

We proceed by using a fixed point theorem. Let B_1 be the space of pairs of continuous functions on $[0, T] \times \mathbb{R}^d$, (Y, Z) so that $\sup_t \|Y(t, \cdot) - \mathrm{Id}\|_\infty \leq 1$ and $\|Z\|_\infty \leq 2\|G_x\|_\infty + 1$. We equip B_1 with the supremum metric.

First we rewrite the system in integral form

$$\begin{aligned}
Y_t(x) &= x + \int_0^t \partial_p H(Y_s(x), Y_s\#\mu, Z_s(x))ds \\
Z_t(x) &= G_x(Y_T(x), Y_T\#\mu) - \int_t^T \partial_x H(Y_s(x), Y_s\#\mu, Z_s(x))ds
\end{aligned}$$

and so we define Φ to be the map on B_1 so that $\Phi(Y, Z)$ is the tuple given by the right hand sides of the above system.

We first show that Φ maps B_1 into itself as long as T is sufficiently small. Fix some

$(Y, Z) \in B_1$ and let $(\tilde{Y}, \tilde{Z}) = \Phi(Y, Z)$. We have

$$\begin{aligned}
\left|\tilde{Z}_t(x)\right| &\leq |G_x(Y_T(x), Y_T\#\mu)| + \int_t^T |\partial_x H(Y_s(x), Y_s\#\mu, Z_s(x))|\, ds \\
&\leq \|G_x\|_\infty + T\sup_s |\partial_x H(Y_s(x), Y_s\#\mu, Z_s(x))| \\
&\leq \|G_x\|_\infty + TC\sup_s (1 + |Z_s(x)|) \\
&\leq \|G_x\|_\infty + TC(2 + 2\|G_x\|_\infty) \\
&\leq (1 + 2TC)\|G_x\|_\infty + 2TC
\end{aligned} \tag{2.6}$$

and so $\|\tilde{Z}\|_\infty \leq 1 + 2\|G_x\|_\infty$ as long as $T \leq \frac{1}{2C}$. Next

$$\left|\tilde{Y}_t(x) - x\right| \leq T\sup_s |\partial_p H(Y_s(x), Y_s\#\mu, Z_s(x))| \leq CT$$

where C is the universal constant that bounds $\partial_p H$ on the set where $p \leq 1 + 2\|G_x\|_\infty$. Hence we have shown that Φ maps B_1 into itself.

Finally we must show that Φ is a contraction. Fix some $(Y, Z), (\tilde{Y}, \tilde{Z}) \in B_1$. Let $(\hat{Y}, \hat{Z}) = \Phi(Y, Z) - \Phi(\tilde{Y}, \tilde{Z})$. We have

$$\begin{aligned}
\left|\hat{Y}_t(x)\right| &\leq \int_0^t \left|\partial_p H(Y_s(x), Y_s\#\mu, Z_s(x)) - \partial_p H(\tilde{Y}_s(x), \tilde{Y}_s\#\mu, \tilde{Z}_s(x))\right| ds \\
&\leq TC\sup_s \left(\left|Y_s(x) - \tilde{Y}_s(x)\right| + \left|Z_s(x) - \tilde{Z}_s(x)\right| + W^2(Y_s\#\mu, \tilde{Y}_s\#\mu)\right) \\
&\leq TC\left(\|Y - \tilde{Y}\|_\infty + \|Z - \tilde{Z}\|_\infty\right)
\end{aligned}$$

at this point we will insist that $T < \frac{1}{8C(C_G+1)}$ so that $\|\hat{Y}\|_\infty \leq \frac{1}{4(C_G+1)}\|(Y, Z) - (\tilde{Y}, \tilde{Z})\|_\infty$. Next

$$\begin{aligned}
\left|\hat{Z}_t(x)\right| \leq &\left|G_x(Y_T(x), Y_T\#\mu) - G_x(\tilde{Y}_T(x), \tilde{Y}_T\#\mu)\right| \\
&+ \int_t^T \left|\partial_x H(Y_s(x), Y_s\#\mu, Z_s(x)) - \partial_x H(\tilde{Y}_s(x), \tilde{Y}_s\#\mu, \tilde{Z}_s(x))\right| ds
\end{aligned}$$

Just as before the integral term is controlled by $TC\|(Y,Z)-(\tilde{Y},\tilde{Z})\|_\infty$. For the G_x term

$$\begin{aligned}
\left|G_x(Y_T(x),Y_T\#\mu)-G_x(\tilde{Y}_T(x),\tilde{Y}_T\#\mu)\right| &\le C_G\left|Y_T(x)-\tilde{Y}_T(x)\right|+C_GW^2(Y_T\#\mu,\tilde{Y}_T\#\mu)\\
&\le C_G\left|Y_T(x)-\tilde{Y}_T(x)\right|+C_G\|Y_T-\tilde{Y}_T\|_\infty\\
&\le 2C_G\|Y_T-\tilde{Y}_T\|_\infty\\
&\le \frac{1}{2}\|(Y,Z)-(\tilde{Y},\tilde{Z})\|_\infty
\end{aligned}$$

This completes the proof that Φ is a contraction.

We now prove the claim without the assumption that $C_G\le\frac{1}{4}$. Let $\tilde{G}=\frac{G}{4C_G}$ and $\tilde{H}(x,\mu,p)=\frac{1}{4C_G}H(x,\mu,4C_Gp)$. Note that $C_{\tilde{G}}=\frac{1}{4}$. Hence by the above the system

$$\begin{aligned}
\dot{\tilde{Y}}_t &= \partial_p\tilde{H}(\tilde{Y}_t,\tilde{Y}_t\#\mu,\tilde{Z}_t)\\
\dot{\tilde{Z}}_t &= -\partial_x\tilde{H}(\tilde{Y}_t,\tilde{Y}_t\#\mu,\tilde{Z}_t)\\
\tilde{Y}_0 &= \mathrm{Id}\\
\tilde{Z}_T &= \partial_x\tilde{G}(\tilde{Y}_T,\tilde{Y}_T\#\mu)
\end{aligned}$$

has a unique solution. It is then easily verified that $(Y_t,Z_t):=(\tilde{Y}_t,4C_G\tilde{Z}_t)$ is the unique solution to our original system.

Finally we have that $\|\tilde{Z}\|_\infty\le 1+2\|\tilde{G}_x\|_\infty$ and so

$$\|Z\|_\infty=4C_G\|\tilde{Z}\|_\infty\le 4C_G+2\|4C_G\tilde{G}_x\|_\infty=4C_G+2\|G_x\|_\infty$$

as desired. □

Remark 2.7.2. It seems likely that one could avoid the rescaling argument by using a more complicated contraction mapping (for example as in [Zha17, Theorem 8.2.1]).

Remark 2.7.3. Note that for Equation 2.6 it was crucial to have that $|\partial_xH(x,\mu,p)|\le C_1(1+|p|)$. In particular a weaker regularity assumption such as saying that ∂_xH is locally bounded in p wouldn't suffice. It is important that the growth is at most linear.

Definition 2.7.4. We let $Y(t,x,\mu), Z(t,x,\mu)$ be the solutions to the system in Proposition 2.7.1.

Lemma 2.7.5. *Y, Z are Lipschitz in t, x and in μ with respect to W_1 where the Lipschitz constant depends only on universal quantities and $\|G_x\|$, $\|G_{xx}\|$, the W_2 Lipschitz constant of G_x.*

Proof. The Lipschitz in the time variable follows directly from the ODE system in Proposition 2.7.1 since $|Z_t|$ is bounded by a universal constant and the derivatives of H are locally bounded in p.

Next we prove Lipschitz in space. This follow a similar proof to Proposition 2.7.1. We first assume that $\|G_{xx}\|_\infty \leq \frac{1}{4}$. Fix some μ, x. Consider the ODE system

$$\dot{A}(t) = \partial_p H(A(t), Y_t\#\mu, B(t))$$
$$\dot{B}(t) = -\partial_x H(A(t), Y_t\#\mu, B(t))$$
$$A(0) = x$$
$$B(T) = \partial_x G(A(T), Y_T\#\mu)$$

to be solved for A, B (of course the solution will be $A, B = Y(\cdot, x, \mu), Z(\cdot, x, \mu)$). We may rewrite the system in the form

$$A(t) = x + \int_0^t \partial_p H(A(s), Y_s\#\mu, B(s))ds$$
$$B(t) = G_x(A(T), Y_T\#\mu) - \int_t^T \partial_x H(A(s), Y_s\#\mu, B(s))ds$$

Let Ψ be the map that sends (A, B) to the right hand side. Just as in the proof to Proposition 2.7.1 we obtain that Ψ is a contraction. Note that if $\tilde{A}, \tilde{B} = Y(\cdot, y, \mu), Z(\cdot, y, \mu)$ then

$$\left|\Psi(\tilde{A}, \tilde{B}) - (\tilde{A}, \tilde{B})\right| = |x - y|$$

and so by the contraction mapping principle we get

$$\left|(A, B) - (\tilde{A}, \tilde{B})\right| \leq C\,|x - y|$$

(here C is a universal constant that is equal to $\frac{1}{1-r}$ where r is the contraction constant of Ψ). Hence Y, Z are Lipschitz in space. By rescaling we can remove the assumption that $\|G_{xx}\|_\infty \leq \frac{1}{4}$.

Now, Lipschitz in the measure variable will also come from the contraction mapping. Again fix μ and let Φ be contraction associated to the initial data μ from the proof of Proposition 2.7.1. Let $\tilde{Y}, \tilde{Z}$ be the solutions with initial data $\tilde{\mu}$ then

$$\left|\Phi(\tilde{Y}, \tilde{Z}) - (\tilde{Y}, \tilde{Z})\right| \leq CW^1(\tilde{Y}_s \# \mu, \tilde{Y}_s \# \tilde{\mu}) \leq CW^1(\mu, \tilde{\mu})$$

and so by the contraction mapping principle we get that

$$\left|(Y, Z) - (\tilde{Y}, \tilde{Z})\right| \leq CW^1(\mu, \tilde{\mu})$$

as desired. □

Lemma 2.7.6. *Suppose that* $\mu = \frac{1}{m+1}\sum_{j=0}^m \delta_{q_i}$ *and* $x \in \{q_1, \ldots, q_m\}$. *Then*

$$\sum_j \left|\partial_{q_j} Y(t, x, \mu)\right| \leq 16$$

if T is sufficiently small.

Proof. Fix some $q \in \mathbb{R}^{(m+1)\times d}$ (labeled $q_0, \ldots, q_k$). Let $\gamma_t^i(q) = Y_t(q_i)$ with the initial condition $\mu = \frac{1}{m+1}\sum_{j=0}^m \delta_{q_i}$. Similarly let $\zeta_t^i(q) = Z_t(q_i)$. Translating the ODE system into our new notation we have

$$\begin{aligned}
\dot{\gamma}_t^i(q) &= H_p(\gamma_t^i(q), \frac{1}{m+1}\sum_{j=0}^m \delta_{\gamma_t^j(q)}, \zeta_t^i(q)) \\
\dot{\zeta}_t^i(q) &= -H_x(\gamma_t^i(q), \frac{1}{m+1}\sum_{j=0}^m \delta_{\gamma_t^j(q)}, \zeta_t^i(q)) \\
\zeta_T^i(q) &= G_x(\gamma_T^i(q), \frac{1}{m+1}\sum_{j=0}^m \delta_{\gamma_T^j(q)}) \\
\gamma_0^i(q) &= q_i
\end{aligned} \tag{2.7}$$

We now differentiate in q_j

$$\partial_{q_j}\dot{\gamma}_t^i(q) = H_{px}\cdot\partial_{q_j}\gamma_t^i(q) + H_{pp}\cdot\partial_{q_j}\zeta_t^i(q) + \frac{1}{m+1}\sum_{k=0}^{m} H_{p\mu}(...,\gamma_t^k(q))\cdot\partial_{q_j}\gamma_t^k(q)$$
$$\partial_{q_j}\dot{\zeta}_t^i(q) = -H_{xx}\cdot\partial_{q_j}\gamma_t^i(q) - H_{xp}\cdot\partial_{q_j}\zeta_t^i(q) - \frac{1}{m+1}\sum_{k=0}^{m} H_{x\mu}(...,\gamma_t^k(q))\cdot\partial_{q_j}\gamma_t^k(q)$$
$$\partial_{q_j}\zeta_T^i(q) = G_{xx}\cdot\partial_{q_j}\gamma_T^i(q) + \frac{1}{m+1}\sum_{k=0}^{m} G_{x\mu}(...,\gamma_t^k(q))\cdot\partial_{q_j}\gamma_T^k(q)$$
$$\partial_{q_j}\gamma_0^i(q) = \delta_{ij}\,\mathrm{Id} \tag{2.8}$$

where we have omitted the argument

$$\gamma_t^i(q), \frac{1}{m+1}\sum_{j=0}^{m}\delta_{\gamma_t^j(q)}, \zeta_t^i(q)$$

from H and $\gamma_t^i(q), \frac{1}{m+1}\sum_{j=0}^{m}\delta_{\gamma_t^j(q)}$ from G.

Note now that since γ, ζ are already known and each ζ^i is bounded by a universal constant, the above is just a finite dimensional system of linear ODE's. Hence we may use Gronwall's inequality to bound $\partial_{q_j}\gamma_t^i, \partial_{q_j}\zeta_t^i$. Fix some q and define

$$\gamma^{i,j} = \sup_t \left|\partial_{q_j}\gamma_t^i(q)\right|$$
$$\zeta^{i,j} = \sup_t \left|\partial_{q_j}\zeta_t^i(q)\right|$$

By applying Gronwall's inequality we obtain

$$\zeta^{i,j} \le C\left(\gamma^{i,j} + \frac{1}{m+1}\sum_{k=0}^{m}\gamma^{k,j} + T\gamma^{i,j} + \frac{T}{m+1}\sum_{k=0}^{m}\gamma^{k,j}\right)e^{CT} \le C\left(\gamma^{i,j} + \frac{1}{m+1}\sum_{k=0}^{m}\gamma^{k,j}\right)$$

and

$$\gamma^{i,j} \le \left(\delta_{ij} + CT\zeta^{i,j} + \frac{CT}{m+1}\sum_{k\in\{0,...,m\},k\neq i}\gamma^{k,j}\right)e^{CT}$$
$$\le \left(\delta_{ij} + CT\left(\gamma^{i,j} + \frac{1}{m+1}\sum_{k=0}^{m}\gamma^{k,j}\right) + \frac{CT}{m+1}\sum_{k\in\{0,...,m\},k\neq i}\gamma^{k,j}\right)e^{CT}$$
$$\le \left(\delta_{ij} + CT\gamma^{i,j} + \frac{CT}{m+1}\sum_{k=0}^{m}\gamma^{k,j}\right)e^{CT}$$

Now suppose that T is so small that $TCe^{TC} \le \frac{1}{4}$. Then

$$\gamma^{i,j} \le 2\delta_{ij} + \frac{1}{2}\gamma^{i,j} + \frac{1}{4(m+1)}\sum_{k=0}^{m}\gamma^{k,j}$$
$$\gamma^{i,j} \le 4\delta_{ij} + \frac{1}{2(m+1)}\sum_{k=0}^{m}\gamma^{k,j}$$

and so summing both sides in i gives $\sum_{k=0}^{m}\gamma^{k,j} \le 8$ and so $\gamma^{i,i} \le 8$ and $\gamma^{i,j} \le \frac{8}{m+1}$ if $i \ne j$. Furthermore we have $\zeta^{i,i} \le C$ and $\zeta^{i,j} \le \frac{C}{m+1}$ if $i \ne j$. □

Lemma 2.7.7. *In the notation of the above lemma we have* $|\partial_t\gamma_t^i|, |\partial_{tt}\gamma_t^i|, |\partial_{tq_i}\gamma_t^i| \le C$. *Furthermore if* $i \ne j$ *then* $\left|\partial_{tq_j}\gamma_t^i\right| \le \frac{C}{m}$. *Furthermore the same bounds hold for* ζ.

Proof. Since ζ_t^i is bounded by a universal constant $|\partial_t\gamma_t^i|, |\partial_t\zeta_t^i| \le C$ follows immediately from the ODE system (2.7).

For the bound on $|\partial_{tt}\gamma_t^i|$ we differentiate (2.7) in time to obtain

$$\partial_{tt}\gamma_t^i = H_{px}\cdot\partial_t\gamma_t^i + H_{pp}\cdot\partial_t\zeta_t^i + \frac{1}{m+1}\sum_{k=0}^{m}H_{p\mu}(..., \gamma_t^k(q))\cdot\partial_t\gamma_t^k(q)$$

where the argument of H is $(\gamma_t^i, \frac{1}{m+1}\sum_{j=0}^{m}\delta_{\gamma_t^j}, \zeta_t^i)$. Taking absolute values of both sides and using the bounds $|\partial_t\gamma_t^i|, |\partial_t\zeta_t^i| \le C$ we see that $|\partial_{tt}\gamma_t^i| \le C$. A similar argument holds for $|\partial_{tt}\zeta_t^i| \le C$.

Finally for the bounds on $\left|\partial_{tq_j}\gamma_t^i\right|$ we recall (2.8). Using $\gamma^{i,i} \le 8$ and $\gamma^{i,j} \le \frac{8}{m+1}$ if $i \ne j$ and the corresponding bounds on ζ we see the claim that $|\partial_{tq_i}\gamma_t^i| \le C$ and $\left|\partial_{tq_j}\gamma_t^i\right| \le \frac{C}{m}$ if $i \ne j$ follows immediately from taking absolute values in the first equation in (2.8). □

Proposition 2.7.8. *We have* $Y, Z \in C^{1,1}$ *jointly in all variables and for each fixed* t, *we have* $Y(t,\cdot), Z(t,\cdot) \in C^2$.

Proof. In light of the above two Lemmas, by [GM22a, Theorems 2.17, 2.19, Corollary 2.18] it suffices to prove the following bounds:

$$\begin{aligned}
\left|\partial_{q_{j_1}q_{j_2}}\gamma^0\right| &\le \frac{C}{m^2}, \quad j_1 \neq j_2 \\
\left|\partial_{q_{j_1}q_{j_1}}\gamma^0\right| &\le \frac{C}{m} \\
\left|\partial_{q_{j_1}q_0}\gamma^0\right| &\le \frac{C}{m} \\
\left|\partial_{q_0q_0}\gamma^0\right| &\le C \\
\left|\partial_{q_{j_1}q_{j_1}q_{j_1}}\gamma^0\right| &\le \frac{C}{m} \\
\left|\partial_{q_{j_1}q_{j_2}q_{j_3}}\gamma^0\right| &\le \frac{C}{m^2}, \quad |\{j_1, j_2, j_3\}| \neq 1
\end{aligned}$$

where $j_a > 0$.

Note that since (2.8) is a finite system of ODE's with a unique bounded solution, the classical ODE theory allows us to differentiate it without needing to check that the differentials exists. We get

$$\begin{aligned}
&\partial_{q_{j_1}q_{j_2}}\dot{\gamma}_t^i = H_{px}\cdot\partial_{q_{j_1}q_{j_2}}\gamma_t^i + H_{pp}\cdot\partial_{q_{j_1}q_{j_2}}\zeta_t^i + \frac{1}{m+1}\sum_{k=0}^m H_{p\mu}\cdot\partial_{q_{j_1}q_{j_2}}\gamma_t^k + \kappa_{1,i,j_1,j_2} \\
&\partial_{q_{j_1}q_{j_2}}\dot{\zeta}_t^i = -H_{xx}\cdot\partial_{q_{j_1}q_{j_2}}\gamma_t^i - H_{xp}\cdot\partial_{q_{j_1}q_{j_2}}\zeta_t^i - \frac{1}{m+1}\sum_{k=0}^m H_{x\mu}\cdot\partial_{q_{j_1}q_{j_2}}\gamma_t^k + \kappa_{2,i,j_1,j_2} \\
&\partial_{q_{j_1}q_{j_2}}\zeta_T^i = G_{xx}\cdot\partial_{q_{j_1}q_{j_2}}\gamma_T^i + \frac{1}{m+1}\sum_{k=0}^m G_{x\mu}\cdot\partial_{q_{j_1}q_{j_2}}\gamma_T^k + \kappa_{3,i,j_1,j_2} \\
&\partial_{q_{j_1}q_{j_2}}\gamma_0^i = 0
\end{aligned}$$

where the κ terms encapsulate all the terms from where the derivative hits H. In particular the κ terms are are quadratic polynomials in $\partial_{q_{j_1}}\gamma_t^i, \partial_{q_{j_2}}\gamma_t^i, \partial_{q_{j_2}}\zeta_t^i, \partial_{q_{j_2}}\zeta_t^i$ with coefficients being derivatives of H, G. Hence using that $\gamma^{i,i} \le C$ and $\gamma^{i,j} \le \frac{C}{m+1}$ if $i \neq j$ from the previous Proposition we will get that $|\kappa_{a,i,j_1,j_2}| \le \frac{C}{m^2}$ if both $j_1, j_2 \neq 0$, $|\kappa_{a,i,j_1,j_2}| \le \frac{C}{m}$ if one of $j_1, j_2 \neq 0$, and $|\kappa_{a,i,j_1,j_2}| \le C$ in any case.

Notice that this system is of the form described in Proposition 2.9.2. Furthermore since ζ^i is already bounded by a universal constant we will obtain that all derivatives of H that appear are bounded by a universal constant. Hence we may apply Proposition 2.9.2 to obtain that

$$\begin{aligned}\left|\partial_{q_{j_1}q_{j_2}}\gamma^0\right| &\leq \frac{C}{m^2}\\ \left|\partial_{q_{j_1}q_0}\gamma^0\right| &\leq \frac{C}{m}\\ \left|\partial_{q_0q_0}\gamma^0\right| &\leq C\end{aligned}$$

when $j_a > 0$.

For the third derivatives terms we differentiate the system yet again:

$$\partial_{q_{j_1}q_{j_2}q_{j_3}}\dot{\gamma}_t^i = H_{px}\cdot\partial_{q_{j_1}q_{j_2}q_{j_3}}\gamma_t^i + H_{pp}\cdot\partial_{q_{j_1}q_{j_2}q_{j_3}}\zeta_t^i + \frac{1}{m+1}\sum_{k=0}^{m} H_{p\mu}\cdot\partial_{q_{j_1}q_{j_2}q_{j_3}}\gamma_t^k + \kappa_{1,i,j_1,j_2,j_3}$$

$$\partial_{q_{j_1}q_{j_2}q_{j_3}}\dot{\zeta}_t^i = -H_{xx}\cdot\partial_{q_{j_1}q_{j_2}q_{j_3}}\gamma_t^i - H_{xp}\cdot\partial_{q_{j_1}q_{j_2}q_{j_3}}\zeta_t^i - \frac{1}{m+1}\sum_{k=0}^{m} H_{x\mu}\cdot\partial_{q_{j_1}q_{j_2}q_{j_3}}\gamma_t^k + \kappa_{2,i,j_1,j_2,j_3}$$

$$\partial_{q_{j_1}q_{j_2}q_{j_3}}\zeta_T^i = G_{xx}\cdot\partial_{q_{j_1}q_{j_2}q_{j_3}}\gamma_T^i + \frac{1}{m+1}\sum_{k=0}^{m} G_{x\mu}\cdot\partial_{q_{j_1}q_{j_2}q_{j_3}}\gamma_T^k + \kappa_{3,i,j_1,j_2,j_3}$$

$$\partial_{q_{j_1}q_{j_2}q_{j_3}}\gamma_0^i = 0$$

where here we will take $j_a > 0$. This time the κ terms are are cubic polynomials in $\partial_{q_{j_1}}\gamma_t^i, \partial_{q_{j_2}}\zeta_t^i$ or quadratic with terms like $\partial_{q_{j_1}}\gamma_t^i\partial_{q_{j_1}q_{j_2}}\gamma_t^i$ or 'summation terms' like

$$\frac{1}{m^3}\sum_{k_1,k_2,k_3}\partial_{q_{j_1}}\gamma_t^{k_1}\partial_{q_{j_2}}\gamma_t^{k_2}\partial_{q_{j_3}}\gamma_t^{k_3}.$$

Since none of the $j_a = 0$ we have that all the first derivative terms are bounded by $\frac{C}{m}$ and all the second derivative terms by $\frac{C}{m^2}$ and so we will get $|\kappa_{b,i,j_1,j_2,j_3}| \leq \frac{C}{m^3}$ and so Proposition 2.9.2 will give that

$$\left|\partial_{q_{j_1}q_{j_2}q_{j_3}}\gamma^0\right| \leq \frac{C}{m^3}.$$

□

Remark 2.7.9. The results of [GM22a] are all local in the sense that they prove $C^{1,1}_{loc}$. For our purpose it is crucial that we have global $C^{1,1}$ bounds (in particular for Proposition 2.9.3).

However reading through the proofs in [GM22a] it seems like the local constants only depend on the local bounds in the assumptions. In particular if we have global bounds then we will get global constants. For example in Theorem [GM22a, Theorem 2.17] reading through the proof it seems like the final "C" depends only on the input $C(K, r)$ and so if the the initial C doesn't depend on K, r then the final C won't either.

2.7.2 Master Equation

We adopt the notation $Y_t(x, \mu, S), Z_t(x, \mu, S)$ to denote the solution to

$$\begin{aligned} \dot{Y}_t &= \partial_p H(Y_t, Y_t \# \mu, Z_t) \\ \dot{Z}_t &= -\partial_x H(Y_t, Y_t \# \mu, Z_t) \\ Y_S &= \mathrm{Id} \\ Z_T &= \partial_x G(Y_T, Y_T \# \mu) \end{aligned} \tag{2.9}$$

Definition 2.7.10. Define V by

$$V(t, x, \mu) = G(Y_T(x, \mu, t), Y_T(\cdot, \mu, t)\#\mu) - \int_t^T L(Y_s(x, \mu, t), \dot{Y}_s(x, \mu, t), Y_s(\cdot, \mu, t)\#\mu) ds$$

This is our candidate solution to the master equation.

Proposition 2.7.11. *V is differentiable in x and $\partial_x V(t, x, \mu) = Z_t(x, \mu, t)$. Furthermore V is Lipschitz in time.*

Proof. Fix some μ, S. Define

$$\begin{aligned} L^{\mu,S}(t, x, v) &= L(x, v, Y_t(\cdot, \mu, S)\#\mu) \\ H^{\mu,S}(t, x, p) &= H(x, p, Y_t(\cdot, \mu, S)\#\mu) \\ G^{\mu,S}(x) &= G(x, Y_T(\cdot, \mu, S)\#\mu) \end{aligned}$$

Consider the optimal control problem

$$\min_\gamma G^{\mu,S}(\gamma(T)) + \int_S^T L^{\mu,S}(s, \gamma(s), \dot{\gamma}(s)) ds$$

Since $Y_t(x,\mu,S), Z_t(x,\mu,S)$ are the unique solutions to the Hamiltonian ODE system (2.9), by [CS04, Theorem 6.3.3] we get that $Y_t(x,\mu,S)$ is the minimizer of the above control problem. Furthermore if we let $\tilde{V}^{\mu,S}(t,x)$ be the associated value function then by [CS04, Theorem 6.4.7] we have that $\tilde{V}^{\mu,S}$ is differentiable at (S,x) for all x. Furthermore by [CS04, Theorem 6.4.8] we have

$$\partial_x \tilde{V}^{\mu,S}(S,x) = Z_S(x,\mu,S)$$

Note that by the definition of the value function

$$\tilde{V}^{\mu,S}(S,x) = G^{\mu,S}(Y_T(x,\mu,S)) + \int_S^T L^{\mu,S}(s, Y_s(x,\mu,S), \dot{Y}_s(x,\mu,S))ds$$

and so $\tilde{V}^{\mu,S}(S,x) = V(S,x,\mu)$.

Next we prove the claim that V is Lipschitz in time. Let $S_1 > S$ be some fixed time. First since $\tilde{V}^{\mu,S}(t,x)$ is the value function associated to an optimal control problem we have that it satisfies the Hamilton Jacobi equation

$$\partial_t \tilde{V}^{\mu,S}(t,x) = -H^{\mu,S}(t,x,\partial_x \tilde{V}^{\mu,S}(t,x)) = -H(x, Z_t(x,\mu,S), Y_t(\cdot,\mu,S)\#\mu)$$

and so $\partial_t \tilde{V}^{\mu,S}(t,x)$ is bounded by a universal constant and so $\tilde{V}^{\mu,S}(t,x)$ is Lipschitz in time.

We have

$$\begin{aligned} |V(S,x,\mu) - V(S_1,x,\mu)| &= \left|\tilde{V}^{\mu,S}(S,x) - V(S_1,x,\mu)\right| \\ &\leq \left|\tilde{V}^{\mu,S}(S_1,x) - V(S_1,x,\mu)\right| + \left|\tilde{V}^{\mu,S}(S,x) - \tilde{V}^{\mu,S}(S_1,x)\right| \\ &\leq |V(S_1,x,Y_{S_1}(\cdot,\mu,S)\#\mu) - V(S_1,x,\mu)| + C\,|S - S_1| \end{aligned}$$

To bound this final expression we first argue that V is Lipschitz in the measure variable with respect to W_1. To see this first recall that from Lemma 2.7.5 we have $Y_t(x,\mu,S)$ is Lipschitz in both x,μ. Hence it follows that the map $\mu \to Y_t(\cdot,\mu,S)\#\mu$ is also Lipschitz with respect to W_1. It now follows directly from the first Hamilton ODE that $\dot{Y}_t(x,\mu,S)$ is

Lipschitz in μ. Hence it follows directly the the definition of the V (Definition 2.7.10) that V is Lipschitz in the measure variable with respect to W_1. Now continuing we have

$$\begin{aligned}|V(S,x,\mu)-V(S_1,x,\mu)| &\leq W_1(Y_{S_1}(\cdot,\mu,S)\#\mu,\mu)+C\,|S-S_1| \\ &\leq \|Y_{S_1}(x,\mu,S)-x\|_{L^\infty(x)}+C\,|S-S_1| \\ &\leq C\,|S-S_1|\end{aligned}$$

where the second inequality is from Lemma 2.4.1 and the last line follows because

$$Y_S(x,\mu,S)=x$$

and Y_t is Lipschitz in time (by Lemma 2.7.5). □

Lemma 2.7.12. *We have*

$$Y_s(Y_t(x,\mu,\eta),Y_t(\cdot,\mu,\eta)\#\mu,t)=Y_s(x,\mu,\eta)$$

when $s\geq t\geq\eta$.

Proof. Fix x,μ,η,t. Let $\tilde{Y}_s$ be the left hand side and $\tilde{Z}_s=Z_s(Y_t(x,\mu,\eta),Y_t(\cdot,\mu,\eta)\#\mu,t)$. Note that $\tilde{Y}_s,\tilde{Z}_s$ solve the ODE

$$\begin{aligned}\dot{\tilde{Y}}_s &= \partial_p H(\tilde{Y}_s,\tilde{Y}_s\#\mu,\tilde{Z}_s) \\ \dot{\tilde{Z}}_s &= -\partial_x H(\tilde{Y}_s,\tilde{Y}_s\#\mu,\tilde{Z}_s) \\ \tilde{Y}_t &= Y_t(x,\mu,\eta) \\ \tilde{Z}_T &= \partial_x G(\tilde{Y}_T,\tilde{Y}_T\#\mu)\end{aligned}$$

which is also solved by $Y_s(x,\mu,\eta),Z_s(x,\mu,\eta)$. Since this ODE has a unique solution by Proposition 2.7.1, we obtain the result. □

Proposition 2.7.13. *V is differentiable in μ. Furthermore $\partial_x V\in C^{1,1}$ jointly in all variables and for each fixed t, we have $\partial_x V(t,\cdot)\in C^2$.*

Proof. For differentiability in μ we consider the two terms in V separately. First consider $I(\mu) := G(Y_T(x,\mu,t), Y_T(\cdot,\mu,t)\#\mu)$ as a function of μ. The lift $\hat{I}$ is given by

$$\hat{I}(A) = \hat{G}(\hat{Y}_T(x,A,t), Y_T(A(\cdot), A\#\mathbb{P}, t)) = \hat{G}(\hat{Y}_T(x,A,t), F(A))$$

where $F(A) := Y_T(A(\cdot), A\#\mathbb{P}, t)$. Now by Proposition 2.9.3 we have that $F(A)$ is Frechet differentiable and $\hat{Y}_T(x,A,t)$ is Frechet differentiable (with respect to A) since Y_T is Wasserstein differentiable from Proposition 2.7.8. Hence $\hat{I}(A)$ is Frechet differentiable and so $I(A)$ is Wasserstein differentiable. A similar argument holds for the integral term in V.

Finally since $\partial_x V(t,x,\mu) = Z_t(x,\mu,t)$ the claims about $\partial_x V(t,x,\mu)$ follow from Proposition 2.7.8. □

Theorem 2.7.14. *V is a solution to the master equation.*

Proof. First $Y_s(Y_t(x,\mu,\eta), Y_t(\cdot,\mu,\eta)\#\mu, t) = Y_s(x,\mu,\eta)$ by Lemma 2.7.12. Hence

$$Y_s\bigg(\cdot, Y_t(\cdot,\mu,\eta)\#\mu, t\bigg)\#\bigg(Y_t(\cdot,\mu,\eta)\#\mu\bigg) = Y_s\bigg(Y_t(\cdot,\mu,\eta), Y_t(\cdot,\mu,\eta)\#\mu, t\bigg)\#\mu = Y_s(\cdot,\mu,\eta)\#\mu$$

and so

$$\begin{aligned}&V(t, Y_t(x,\mu,\eta), Y_t(\cdot,\mu,\eta)\#\mu)\\ &= G(Y_T(x,\mu,\eta), Y_T(\cdot,\mu,\eta)\#\mu) - \int_t^T L(Y_s(x,\mu,\eta), \dot{Y}_s(x,\mu,\eta), Y_s(\cdot,\mu,\eta)\#\mu)ds\end{aligned}$$

We may now differentiate both sides with respect to t for almost every t to get

$$\partial_t V + \partial_x V \cdot \dot{Y}_t + \int \partial_\mu V(..., \tilde{x}) \cdot \dot{Y}_t(\tilde{x},\mu,\eta) dY_t\#\mu = L(Y_t, \dot{Y}_t, Y_t\#\mu)$$

where the argument of Y_t is (x,μ,η) and the argument of V is $(t, Y_t, Y_t\#\mu)$. When $\eta = t$ we get $Y_t(x,\mu,t) = x$ and so evaluating the above when $\eta = t$ we get

$$\partial_t V(t,x,\mu) + \partial_x V(t,x,\mu) \cdot \dot{Y}_t(x,\mu,t) + \int \partial_\mu V(t,x,\mu,\tilde{x}) \cdot \dot{Y}_t(\tilde{x},\mu,t) d\mu = L(x, \dot{Y}_t(x,\mu,t), \mu)$$

Now by the Hamiltonian ODE's we have

$$\begin{aligned}\dot{Y}_t(x,\mu,t) &= \partial_p H(Y_t(x,\mu,t), Y_t(\cdot,\mu,t)\#\mu, Z_t(x,\mu,t)) \\ &= \partial_p H(x,\mu,Z_t(x,\mu,t)) = \partial_p H(x,\mu,\partial_x V(t,x,\mu))\end{aligned}$$

where the last equality is from Proposition 2.7.11. Plugging in we get

$$\begin{aligned}&\partial_t V(t,x,\mu) + \partial_x V(t,x,\mu)\cdot\partial_p H(x,\mu,\partial_x V(t,x,\mu)) - L(x,\partial_p H(x,\mu,\partial_x V(t,x,\mu)),\mu) \\ &+ \int \partial_\mu V(t,x,\mu,\tilde{x})\cdot\partial_p H(\tilde{x},\mu,\partial_x V(t,\tilde{x},\mu))d\mu(\tilde{x}) = 0\end{aligned}$$

Using that $H = L^*$ we have

$$\partial_x V(t,x,\mu)\cdot\partial_p H(x,\mu,\partial_x V(t,x,\mu)) - L(x,\partial_p H(x,\mu,\partial_x V(t,x,\mu)),\mu) = H(x,\mu,\partial_x V(t,x,\mu))$$

and so we get

$$\partial_t V(t,x,\mu) + H(x,\mu,\partial_x V(t,x,\mu)) + \int \partial_\mu V(t,x,\mu,\tilde{x})\cdot\partial_p H(\tilde{x},\mu,\partial_x V(t,\tilde{x},\mu))d\mu(\tilde{x}) = 0$$

as desired.

We remark that since the terms besides $\partial_t V(t,x,\mu)$ are continuous in time we get that V was actually continuously differentiable in time as opposed to only almost everywhere. □

2.8 Well-Posedness for the Master Equation

With the a-priori estimates in hand it is now standard to obtain well-posedness for the master equation. For convenience of the reader we include the full proof.

Theorem 2.8.1. *There is a unique classical solution to the master equation.*

Proof. We prove uniqueness first. Suppose that $V, \tilde{V}$ are two classical solutions to the master equation. Let

$$T_1 = \inf\{t : V(s,x,\mu) = \tilde{V}(s,x,\mu) \quad \forall s \in [t,T]\}$$

Since $V, \tilde{V}$ are continuous we have $V(T_1, \cdot) = \tilde{V}(T_1, \cdot)$. We will assume without loss of generality that $T_1 = T$.

Let δ be the constant from Proposition 2.7.1. Choose some fixed s, y, μ with $s \in (T-\delta, T)$ so that $V(s, y, \mu) \neq \tilde{V}(s, y, \mu)$. We may shift our $V, \tilde{V}$ in time by s to reduce to the situation where $V(0, y, \mu) \neq \tilde{V}(0, y, \mu)$ and $T < \delta$.

Let $Y_t : \mathbb{R}^d \to \mathbb{R}^d$ be the solution to

$$\dot{Y}_t = \partial_p H(Y_t, Y_t \# \mu, \partial_x V(t, Y_t, Y_t \# \mu))$$
$$Y_0 = \mathrm{Id}$$

and define $Z_t(x) = \partial_x V(t, Y_t(x), Y_t \# \mu)$. We claim that

$$\dot{Z}_t(x) = -\partial_x H(Y_t(x), Y_t \# \mu, Z_t(x))$$

Indeed if we define $U(t, x) = V(t, x, Y_t \# \mu)$ then U is the value function for an optimal control problem with Hamiltonian $H(\cdot, Y_t \# \mu, \cdot)$ (this follows from the exact same reasoning as Lemma 2.5.1). Furthermore, we see that $\partial_x U(t, x) = \partial_x V(t, x, Y_t \# \mu)$. From

$$\dot{Y}_t = \partial_p H(Y_t, Y_t \# \mu, \partial_x U(t, Y_t))$$

we see that $Y_t(x)$ are actually the optimal paths for the optimal control problem. Hence it follows from Hamilton's ODE's that

$$\dot{Z}_t(x) = -\partial_x H(Y_t(x), Y_t \# \mu, Z_t(x))$$

since $Z_t(x) = \partial_x U(t, Y_t(x))$. Finally we remark that since $V(T, \cdot) = G(\cdot)$ we have

$$Z_T = \partial_x G(Y_T, Y_T \# \mu).$$

We now repeat the exact same procedure to define $\tilde{Y}_t$, $\tilde{Z}_t$, and $\tilde{U}$ only we use $\tilde{V}$ in place of V. By the uniqueness from Proposition 2.7.1 we get that $Y_t = \tilde{Y}_t$ and $Z_t = \tilde{Z}_t$. In particular

$Y_t\#\mu = \tilde{Y}_t\#\mu$ and so the optimal control problems associated to $U, \tilde{U}$ are the same. Hence $U = \tilde{U}$. In particular

$$V(0, y, \mu) = U(0, y) = \tilde{U}(0, y) = \tilde{V}(0, y, \mu)$$

as desired.

Now we prove existence. We proceed by contradiction. Suppose $T_1 > 0$ is the smallest time so that there is a classical solution to the master equation, V, on $(T_1, T]$ so that $\partial_x V \in C^2$.

Let δ be the constant from Proposition 2.7.1 except that we take the a-priori bounds from Theorems 2.5.4 and 2.6.1 and Corollary 2.5.5 in place of $\|G_x\|_\infty, \|G_{xx}\|_\infty$ and the W_2 Lipschitz constant of G_x.

Set $T_2 = T_1 + \frac{\delta}{2}$ and $G_2 = V(T_2, \cdot)$. By Theorem 2.4.4, V is displacement monotone and so we may use Theorems 2.5.4 and 2.6.1 and Corollary 2.5.5 to bound $\|\partial_x G_2\|_\infty, \|\partial_{xx} G_2\|_\infty$ and the W_2 Lipschitz constant of $\partial_x G_2$.

Now we use Theorem 2.7.14 to construct a classical solution V_2 to the master equation on $(T_2 - \delta, T_2) = (T_1 - \frac{\delta}{2}, T_1 + \frac{\delta}{2})$ so that $V_2(T_1 + \frac{\delta}{2}, \cdot) = G_2(\cdot) = V(T_1 + \frac{\delta}{2})$. By the uniqueness result above we get $V_2 = V$ for all $t \in (T_1, T_1 + \frac{\delta}{2})$. Hence we may use V_2 to extend our solution V to a solution to the master equation on $(T_1 - \frac{\delta}{2}, T]$. This contradicts T_1 being the smallest time. $\square$

2.9 Technical Lemmas

2.9.1 Forward Backward ODE Systems

We first show Existence/Uniqueness for Linear Forward Backward ODE's. These results are needed for our method to obtain well-posedness for the deterministic master equation. While they will not be surprising to experts we are not aware of any reference for these exact results.

Lemma 2.9.1. *The Linear Forward Backward ODE*

$$\begin{aligned}
\dot{\gamma}_t^i &= \sum_j A_1^{ij}(t)\gamma_t^j + \sum_j B_1^{ij}(t)\zeta_t^j + K_1(t) \\
\dot{\zeta}_t^i &= \sum_j A_2^{ij}(t)\gamma_t^j + \sum_j B_2^{ij}(t)\zeta_t^j + K_2(t) \\
\gamma_0^i &= q_i \\
\zeta_T^i &= \sum_j G^{ij}\gamma_T^j + K_3
\end{aligned}$$

has a unique solution as long as $T < \frac{1}{8}\left(\sum_j \left|A_1^{ij}(s)\right| + \left|B_1^{ij}(s)\right|\right)^{-1}$.

Proof. We first prove this under the additional assumption that $\sum_j |G^{ij}| \leq \frac{1}{2}$.

We proceed by contraction mapping principle. Rewrite the ODE in the integral form

$$\begin{aligned}
\gamma_t^i &= q_i + \int_0^t \sum_j A_1^{ij}(s)\gamma_s^j + \sum_j B_1^{ij}(s)\zeta_s^j + K_1(s)ds \\
\zeta_t^i &= \sum_j G^{ij}\gamma_T^j + K_3 + \int_t^T \sum_j A_2^{ij}(s)\gamma_s^j + \sum_j B_2^{ij}(s)\zeta_s^j + K_2(s)ds
\end{aligned}$$

Let B be that space of all continuous (γ_t^i, ζ_t^i) such that $\|\gamma_t^i\|_\infty, \|\zeta_t^i\|_\infty < \infty$ where the L^∞ norm is taken over time. We equip B with the supremum metric.

Let Φ be the map that sends (γ_t^i, ζ_t^i) to the right hand side of the above system. We see that Φ maps B to B. We show that Φ is a contraction. We proceed by direct computation.

Let $(\hat{\gamma}_t^i, \hat{\zeta}_t^i) = \Phi(\gamma_t^i, \zeta_t^i) - \Phi(\tilde{\gamma}_t^i, \tilde{\zeta}_t^i)$. Set $\eta = \|(\gamma_t^i, \zeta_t^i) - (\tilde{\gamma}_t^i, \tilde{\zeta}_t^i)\|_\infty$ We see that

$$\left|\hat{\gamma}_t^i\right| \leq \int_0^t \sum_j \left|A_1^{ij}(s)\right| \left|\gamma_s^j - \tilde{\gamma}_s^j\right| + \sum_j \left|B_1^{ij}(s)\right| \left|\zeta_s^j - \tilde{\zeta}_s^j\right| ds \leq T\eta \left(\sum_j \left|A_1^{ij}(s)\right| + \left|B_1^{ij}(s)\right|\right)$$

and

$$
\begin{aligned}
\left|\hat{\zeta}_t^i\right| &\le \int_0^t \sum_j \left|A_2^{ij}(s)\right| \left|\gamma_s^j - \tilde{\gamma}_s^j\right| + \sum_j \left|B_2^{ij}(s)\right| \left|\zeta_s^j - \tilde{\zeta}_s^j\right| ds + \sum_j \left|G^{ij}\right| \left|\gamma_T^j - \tilde{\gamma}_T^j\right| \\
&\le \eta \left(T \sum_j \left|A_2^{ij}(s)\right| + T \sum_j \left|B_2^{ij}(s)\right| + \sum_j \left|G^{ij}\right| \right) \\
&\le \eta \left(T \sum_j \left|A_2^{ij}(s)\right| + T \sum_j \left|B_2^{ij}(s)\right| + \frac{1}{2} \right)
\end{aligned}
$$

and so Φ is a contraction as long as $T < \frac{1}{4}\left(\sum_j \left|A_1^{ij}(s)\right| + \left|B_1^{ij}(s)\right|\right)^{-1}$.

Now we consider the case where $K_G := \sum_j |G^{ij}| \ge \frac{1}{2}$. Consider the rescaled system

$$
\begin{aligned}
\dot{\tilde{\gamma}}_t^i &= \sum_j \frac{A_1^{ij}(t)}{K_G} \tilde{\gamma}_t^j + \sum_j B_1^{ij}(t) \tilde{\zeta}_t^j + \frac{K_1(t)}{K_G} \\
\dot{\tilde{\zeta}}_t^i &= \sum_j \frac{A_2^{ij}(t)}{K_G} \tilde{\gamma}_t^j + \sum_j B_2^{ij}(t) \tilde{\zeta}_t^j + \frac{K_2(t)}{K_G} \\
\tilde{\gamma}_0^i &= q_i \\
\tilde{\zeta}_T^i &= \sum_j \frac{G^{ij}}{K_G} \tilde{\gamma}_T^j + \frac{K_3}{K_G}
\end{aligned}
$$

This system satisfies the assumptions of our contraction mapping proof and so it has a unique solution $(\tilde{\gamma}_t^i, \tilde{\zeta}_t^i)$. We then see that $(\gamma_t^i, \zeta_t^i) := (\tilde{\gamma}_t^i, K_G \tilde{\zeta}_t^i)$ is the solution the the original (un-rescaled) system. □

Proposition 2.9.2. *Consider the following linear forward backward ODE*

$$
\begin{aligned}
\dot{\gamma}_{t,j}^i &= H_{1,1}(t) \cdot \gamma_{t,j}^i + H_{1,2}(t) \cdot \zeta_{t,j}^i + \sum_{k \ne i} H_{2,k}(t) \cdot \gamma_{t,j}^k + \kappa_{1,i,j}(t) \\
\dot{\zeta}_{t,j}^i &= H_{3,1}(t) \cdot \gamma_{t,j}^i + H_{3,2}(t) \cdot \zeta_{t,j}^i + \sum_{k \ne i} H_{4,k}(t) \cdot \gamma_{t,j}^k + \kappa_{2,i,j}(t) \\
\gamma_{0,j}^i &= 0 \\
\zeta_{T,j}^i &= H_{5,1} \cdot \gamma_{T,j}^i + \sum_{k \ne i} H_{6,k} \cdot \gamma_{T,j}^k + \kappa_{3,i,j}
\end{aligned}
$$

Here i runs from $0,\dots,m$ and j runs over an arbitrary index set (typically $\{0,\dots,m\}^a$).

Define $H^{a,b} = \sup_t |H_{a,b}(t)|$, $\kappa^{a,b,c} = \sup_t |\kappa_{a,b,c}(t)|$, $\gamma^{a,b} = \sup_t \left|\gamma^a_{t,b}\right|$, and $\zeta^{a,b} = \sup_t \left|\zeta^a_{t,b}\right|$. Let $\kappa^{i,j} = \sum_{a=1,2,3} \kappa^{a,i,j}$.

Suppose that there is a constant C so that

$$\left|H^{1,1}\right|, \left|H^{1,2}\right|, \left|H^{3,1}\right|, \left|H^{3,2}\right|, \left|H^{5,1}\right| \leq C$$

and $\left|H^{a,k}\right| \leq \frac{C}{m}$ for $a \in \{2,4,6\}$ and $k \in \{0,\dots,m\}$.

Then if T is sufficiently small (which depends only on C) we have

$$\gamma^{i,j} \leq \kappa^{i,j} + \frac{1}{m}\sum_k \kappa^{k,j}$$

and

$$\zeta^{i,j} \leq C\left(\kappa^{i,j} + \frac{1}{m}\sum_k \kappa^{k,j}\right).$$

Proof. To simplify the notation we will drop the $\cdot$ used for dot products. All sums over k are over $k \in \{0,\dots,m\}$ but $k \neq i$.

By Gronwall's inequality we have

$$\begin{aligned}\zeta^{i,j} &\leq 2\left(\kappa^{3,i,j} + H^{5,1}\gamma^{i,j} + \sum_k H^{6,k}\gamma^{k,j} + T\left(\kappa^{2,i,j} + \sum_k H^{4,k}\gamma^{k,j} + H^{3,1}\gamma^{i,j}\right)\right)e^{TH^{3,2}} \\ &\leq C\left(\kappa^{3,i,j} + \gamma^{i,j} + \kappa^{2,i,j} + \frac{1}{m}\sum_k \gamma^{k,j}\right)\end{aligned}$$

and

$$\gamma^{i,j} \leq T\left(\kappa^{1,i,j} + H^{1,2}\zeta^{i,j} + \sum_k H^{2,k}\gamma^{k,j}\right)e^{TH^{1,1}} \leq CT\left(\kappa^{1,i,j} + \zeta^{i,j} + \frac{1}{m}\sum_k \gamma^{k,j}\right)$$

Combining these two we get

$$\begin{aligned}\gamma^{i,j} &\leq CT\left(\kappa^{1,i,j} + \kappa^{3,i,j} + \gamma^{i,j} + \kappa^{2,i,j} + \frac{1}{m}\sum_k \gamma^{k,j} + \frac{1}{m}\sum_k \gamma^{k,j}\right) \\ &\leq CT\left(\gamma^{i,j} + \frac{1}{m}\sum_k \gamma^{k,j} + \kappa^{i,j}\right)\end{aligned}$$

Now if $CT \leq \frac{1}{2}$ we get

$$\gamma^{i,j} \leq 2CT \left(\frac{1}{m} \sum_k \gamma^{k,j} + \kappa^{i,j} \right)$$

Summing over i on both sides we get

$$\sum_{i=0}^{m} \gamma^{i,j} \leq 2CT \left(\sum_k \gamma^{k,j} + \sum_{i=0}^{m} \kappa^{i,j} \right)$$

and so if $2CT \leq \frac{1}{2}$ we have

$$\sum_{i=0}^{m} \gamma^{i,j} \leq 4CT \sum_{i=0}^{m} \kappa^{i,j}$$

which yields

$$\gamma^{i,j} \leq \kappa^{i,j} + \frac{1}{m} \sum_k \kappa^{k,j}$$

as desired. Next plugging into our bound on $\zeta^{i,j}$ we get

$$\zeta^{i,j} \leq C \left(\kappa^{3,i,j} + \kappa^{i,j} + \frac{1}{m} \sum_k \kappa^{k,j} + \kappa^{2,i,j} + \frac{1}{m} \sum_{i=0}^{m} \kappa^{i,j} \right) \leq C \left(\kappa^{i,j} + \frac{1}{m} \sum_{i=0}^{m} \kappa^{i,j} \right)$$

as desired. □

Proposition 2.9.3. *Suppose that* $Y : \mathbb{R}^d \times \mathcal{P}_2 \to \mathbb{R}^d$ *is* $C^{1,1}$ *(in the sense of [GM22a, Theorem 2.17]). Then the map* $F : L^2(\Omega) \to L^2(\Omega)$ *defined by*

$$F(X)[\omega] = Y(X(\omega), X\#\mathbb{P})$$

is Frechet differentiable and the Frechet derivative is Lipschitz.

Proof. We let $\hat{Y} : \mathbb{R}^d \times L^2(\Omega) \to \mathbb{R}^d$ be the lift of Y defined by $\hat{Y}(x, A) = Y(x, A\#\mathbb{P})$. We have

$$\hat{Y}(A(\omega) + B(\omega), A + B) = \hat{Y}(A(\omega), A + B) + D_x\hat{Y}(A(\omega), A + B) \cdot B(\omega) + O(|B(\omega)|^2)$$

where the constant in the big O doesn't depend on A, B, ω. Hence we obtain

$$\|\hat{Y}(A(\cdot)+B(\cdot), A+B)-\hat{Y}(A(\cdot), A+B)-D_x\hat{Y}(A(\cdot), A+B)\cdot B(\cdot)\|_2 \leq C\||B(\cdot)|^2\|_2 \leq C\|B\|_2^2 \tag{2.10}$$

Next because $D_x\hat{Y}$ is Lipschitz we have that

$$\left|D_x\hat{Y}(A(\omega), A+B) - D_x\hat{Y}(A(\omega), A)\right| \leq C\|B\|_2$$

for each fixed ω. Hence

$$\left|D_x\hat{Y}(A(\omega), A+B)\cdot B(\omega) - D_x\hat{Y}(A(\omega), A)\cdot B(\omega)\right| \leq C\|B\|_2\,|B(\omega)|$$

and so

$$\|D_x\hat{Y}(A(\cdot), A+B)\cdot B(\cdot) - D_x\hat{Y}(A(\cdot), A)\cdot B(\cdot)\|_2 \leq C\|B\|_2^2$$

Plugging into (2.10) we get

$$\|\hat{Y}(A(\cdot)+B(\cdot), A+B) - \hat{Y}(A(\cdot), A+B) - D_x\hat{Y}(A(\cdot), A)\cdot B(\cdot)\|_2 \leq C\|B\|_2^2$$

Because Y is differentiable in the measure variable we have that $\hat{Y}$ is Frechet differentiable in its second variable. We denote this derivative by $D_\mu\hat{Y}$. For a fixed $x \in \mathbb{R}^d$ we have $D_\mu\hat{Y}(x,\cdot) : L^2(\Omega) \to (L^2(\Omega) \to \mathbb{R}^d)$. Hence we use the notation $D_\mu\hat{Y}(x, A)[\cdot] : L^2(\Omega) \to \mathbb{R}^d$.

Next we have for each fixed ω

$$\hat{Y}(A(\omega), A+B) = \hat{Y}(A(\omega), A) + D_\mu\hat{Y}(A(\omega), A)[B] + O(\|B\|_2^2)$$

and so

$$\|\hat{Y}(A(\cdot), A+B) - \hat{Y}(A(\cdot), A) - D_\mu\hat{Y}(A(\cdot), A)[B]\|_2 \leq C\|B\|_2^2$$

Again plugging in we get

$$\|\hat{Y}(A(\cdot)+B(\cdot), A+B) - \hat{Y}(A(\cdot), A) - D_\mu\hat{Y}(A(\cdot), A)[B] - D_x\hat{Y}(A(\cdot), A)\cdot B(\cdot)\|_2 \leq C\|B\|_2^2$$

Finally recall that

$$F(X)[\omega] = Y(X(\omega), X\#\mathbb{P}) = \hat{Y}(X(\omega), X)$$

and so

$$\|F(A+B)[\cdot] - F(A)[\cdot] - D_\mu \hat{Y}(A(\cdot), A)[B] - D_x \hat{Y}(A(\cdot), A) \cdot B(\cdot)\|_2 \leq C\|B\|_2^2$$

as desired. $\square$

CHAPTER 3

Common Noise Master Equation

3.1 Introduction

Master equations are PDEs of hyperbolic type, whose solutions depend both on the state of individual agents (typically a variable in a finite dimensional Euclidean space) and on the agents' distribution (typically a Borel probability measure supported over the state space of the agents). Beside their independent interest, one of the main motivations for studying these equations lies in the fact that their classical solutions can be used to provide quantitative rates of convergence for the closed loop Nash equilibria of stochastic differential games, when the number of agents tends to infinity (cf. [CDLL19]). They can serve also as important tools in showing large deviation principles, concentration of measure and central limits theorems for these games (see [DLR19, DLR20]).

Because of the infinite dimensional character of these equations, their well-posedness provide great mathematical challenges and so their investigation has gained considerable attention in the community in the past decade. Classical solutions to the master equation are known to exist under certain assumptions on the data, which are responsible for the uniqueness of the MFG Nash equilibria of the underlying game. These assumptions can be roughly grouped into two categories: (i) the data satisfy some sort of smallness condition (related to the time horizon, to the Hamiltonian, to a specific subclass of probability measures, etc.) or (ii) the data fulfill suitable monotonicity conditions.

In the case (i), besides the smallness assumption typically there is no need to impose additional structural assumptions on the data (such as separability or convexity of the underlying Hamiltonian or final datum or the presence of a non-degenerate idiosyncratic noise) governing the game (see for instance [CD18b, GS15, May20, AM23]). The question regarding the global well-posedness of master equations (in the class of classical solutions) is way more subtle and this is understood in suitably defined monotone regimes (cf. case (ii)). In the literature to date, these are essentially classified in two major groups: the so-called Lasry–Lions monotonicity and displacement monotonicity conditions, which are in general in dichotomy with each other. Historically, the Lasry–Lions monotonicity condition was used first for the global well-posedness of the master equation (see [CCD22, CDLL19, CD18b]). When dealing with classical solutions, it worth mentioning that the Lasry–Lions monotonicity condition on its own is in general not enough for the global well-posedness of the underlying master equation, unless a non-degenerate idiosyncratic noise (or stronger convexity assumption on the data) is also present and the corresponding Hamiltonians are separable in the momentum and measure variables, i.e. they possess a decomposition of the form

$$H(x, \mu, p) := H_0(x, p) - F(x, \mu), \tag{3.1.1}$$

for some $H_0 : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and $F : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ (where the state space of the agents is $\mathbb{R}^d$ and $\mathscr{P}_2(\mathbb{R}^d)$ stands for the set of Borel probability measures with finite second moment, supported on $\mathbb{R}^d$, describing the agents' distribution).

Displacement monotonicity (which stems from the notion of displacement convexity arising in optimal transport, [McC97]) is an alternative condition which guarantees the existence and uniqueness of classical solutions to the master equation. Prior to using this condition in the context of master equations, under different names (as 'weak monotonicity' or 'L-monotonicity') this condition has already appeared in works on MFG (see [Ahu16, ARY19] and [CD18a, Section 3.4.3]) and on FBSDEs of McKean–Vlasov type (see [CD15]). This condition turned out to be sufficient in the case of deterministic potential master equations in the lack of the regularizing effect of the idiosyncratic noise ([GM22a, BGY24]), or for a general class of non-separable Hamiltonians in the presence of non-degenerate idiosyncratic noise ([GMMZ22]). In the presence of Lasry–Lions monotonicity and non-degenerate idiosyncratic noise, for separable Hamiltonians suitable notions of weak solutions have been proposed if the data are not regular enough ([Ber21, MZ22a]). In such cases, however, we still have uniqueness of the MFG Nash equilibria. The recent work [GM22b] proposed a new notion of monotonicity condition, which is also a sufficient condition for the global well-posedness of the master equation. This condition is in general in dichotomy with both Lasry–Lions and displacement monotonicity conditions (see [GM23]).

When monotonicity conditions are violated and the uniqueness of the MFG Nash equilibria does not hold, the classical well-posedness theory for the master equation breaks down in finite time and it is a great challenge to define suitable notions of weak solutions, which may help selecting specific equilibria of the game. In this direction it worth mentioning the recent breakthrough [CD24] which proposes a notion of weak solution (in the spirit of entropy solutions) for potential MFG master equations in the presence of non-degenerate idiosyncratic noise. It has been pointed out in [GM22b] that weak solutions in entropy sense (although different from the ones in [CD24]) might in general not select MFG Nash equilibria of the underlying game.

As discussed above, the Lasry–Lions monotonicity condition on the data without the presence of a non-degenerate idiosyncratic noise in general cannot guarantee the uniqueness of solutions to the MFG system (see the discussion in [GM23]) and so, the existence of a classical solution to the corresponding master equation. In the lack of non-degenerate idiosyncratic noise, the literature discusses two important classes of examples: purely deterministic problems and problems driven by a common noise. In the case of deterministic Lasry–Lions monotone MFG systems, [CP20a, Theorem 1.8] presents a uniqueness result under the additional assumption that the measure component is essentially bounded.

MFG systems and master equations driven by common noise only or in presence of common noise and degenerate idiosyncratic noise have been recently investigated in the series of interesting works [CS22a, CS22b, CSS22]. In such cases, a notion of weak solution for the master equation is obtained in the spirit of monotone solutions proposed in [Ber21]. The study of MFG with common noise goes back to the works [Ahu16, CDL16]. Interestingly, already in these early works it has been discussed that additional convexity properties of the value function can render a stronger notion of solutions to MFG with common noise (than the ones in [CS22b, CSS22]). It is well-known now that the convexity of the value function in the state variable is strongly linked to the displacement monotonicity of the data (see [GM22a, GMMZ22, MM24]).

To the best of our knowledge, there are only very few works in the literature studying the global existence and uniqueness of classical solutions to the master equation in lack of non-degenerate idiosyncratic noise: [GM22a] considers potential deterministic master equation in the case of separable Hamiltonians and displacement convexity; [GM22b] studies a class of deterministic master equations in the presence of a different monotonicity condition; a particular dimension reduction technique and the associated monotonicity conditions allowed the authors of [LLS22] to obtain global classical solutions to the deterministic master equation. Finally, in [CS22a] the authors obtain weak monotone solutions to a class of time independent master equations both in the deterministic setting and driven by a common noise.

Our objective in this chapter is to show the global existence and uniqueness of classical solutions to the master equation in the lack of idiosyncratic noise and the presence of displacement monotone data. Our result cover both the deterministic problem and the one driven purely by a common noise. In this chapter we consider the master equation (1.1) with $\beta = 0$ but $\beta_0 \neq 0$. For the convenience of the reader we reproduce it here:

$$\begin{cases} -\partial_t V(t,x,\mu) + H(x,\mu,\partial_x V) - (\mathcal{N}V)(t,x,\mu) - \frac{\beta^2}{2}\Delta_{com} V(t,x,\mu) = 0, \\ \qquad\qquad in\ (0,T)\times\mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d), \\ V(T,x,\mu) = G(x,\mu), in\ \mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d) \end{cases}$$

where

$$(\mathcal{N}V)(t,x,\mu) := -\int_{\mathbb{R}^d} \partial_\mu V(t,x,\mu,\tilde{x}) \cdot \partial_p H(\tilde{x},\mu,\partial_x V(t,\tilde{x},\mu))\, d\mu(\tilde{x})$$

and

$$\begin{aligned}\Delta_{com} V(t,x,\mu) := &\operatorname{tr}(\partial_{xx} V(t,x,\mu)) + \int_{\mathbb{R}^d} \operatorname{tr}(\partial_{\tilde{x}\mu} V(t,x,\mu,\tilde{x}))\, d\mu(\tilde{x}) \\ &+ 2\int_{\mathbb{R}^d} \operatorname{tr}(\partial_{x\mu} V(t,x,\mu,\tilde{x}))\, d\mu(\tilde{x}) \\ &+ \iint_{\mathbb{R}^d\times\mathbb{R}^d} \operatorname{tr}(\partial_{\mu\mu} V(t,x,\mu,\tilde{x},\bar{x}))\, d\mu(\tilde{x})\, d\mu(\bar{x}).\end{aligned}$$

Here $T > 0$ is the time horizon of the game, $\beta \geq 0$ stands for the intensity of the common noise represented by a Brownian motion $(B_t^0)_{t\in[0,T]}$ on $\mathbb{R}^d$, $H : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ and $G : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ are the Hamiltonian and the final cost function, respectively.

Definition 3.1.1. A function $V : (0,T) \times \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is said to be a classical solution to the master equation if all of the derivatives that appear in the equation exist and are continuous (with respect to Euclidean distance and W_1) and V satisfies the master equation pointwise.

The master equation (3.1.2) is strongly linked to the following mean field games system

$$\begin{cases} du(t,x) &= -\Big[\operatorname{tr}\big(\frac{\beta^2}{2}\partial_{xx}u(t,x) + \beta\partial_x v^\top(t,x)\big) - H(x,\rho(t,\cdot),\partial_x u(t,x))\Big]dt \\ & \quad +\beta v(t,x)\cdot dB_t^0, in\ (t_0,T)\times\mathbb{R}^d, \\ d\rho(t,x) &= \Big[\frac{\beta^2}{2}\operatorname{tr}\big(\partial_{xx}\rho(t,x)\big) + div(\rho(t,x)\partial_p H(x,\rho(t,\cdot),\partial_x u(t,x)))\Big]dt \\ & \quad -\beta\partial_x\rho(t,x)\cdot dB_t^0, in\ (t_0,T)\times\mathbb{R}^d, \\ & \quad \rho(t_0,\cdot) = \mu,\ u(T,\cdot) = G(\cdot,\rho(T,\cdot)), \\ & \quad in\ \mathbb{R}^d. \end{cases} \tag{3.1.2}$$

The solution to (3.1.2) is a triple (ρ,u,v), $\mathbb{F}^0$-progressively measurable, which serves formally as the system of generalized characteristics for (3.1.2). We note that if $\beta > 0$, then $\rho(t,\cdot,\omega)$

is a random probability measure. Conversely, the solution V to the master equation (3.1.2) also serves as the decoupling field for this forward-backward PDE system, i.e.

$$u(t, x, \omega) = V(t, x, \rho(t, \cdot, \omega)). \tag{3.1.3}$$

The description of results. As our main result (Theorem 3.5.3) we show the global in time existence and uniqueness of a classical solution to (3.1.2), by assuming that H and G satisfy suitable displacement monotonicity (and regularity) conditions. The displacement monotonicity condition on H and G are the same as the ones proposed in [GMMZ22, Definition 2.2(ii), Definition 3.4]. This results can be seen as a completion of the program initiated in [GM22a, GMMZ22]. The roadmap to the proof of our main result is similar in spirit to the one used in [GMMZ22], but several new ideas were necessary to fulfil this because of the lack of the idiosyncratic noise.

Let us discuss the main similarities and differences in the two approaches. First, the heart of our analysis is the a priori propagation of the displacement monotonicity: if V is a classical solution to (3.1.2) and H and G are displacement monotone, so is $V(t, \cdot, \cdot)$. Displacement monotonicity will readily imply that $\partial_x V(t, x, \cdot)$ is Lipschitz continuous with respect to the metric W_2 (with a Lipschitz constant depending on the data and on $\|\partial_{xx} V\|_{L^\infty}$). These two properties follow similarly as in [GMMZ22]. It is well-known (see [CD18b]) that the master equation is well-posed for short time if the data is regular enough (without any monotonicity assumptions). The short time horizon depends on the Lipschitz constant of $\partial_x V(t, \cdot, \cdot)$ (in the metric W_1 for the measure variable). To show that this Lipschitz constant is a priori bounded, we use two arguments. First, the uniform a priori estimates on $\|\partial_{xx} V\|_{L^\infty}$ are a consequence of the semi-concavity bounds (a result of classical optimal control arguments) and convexity (a consequence of the displacement monotonicity) of $V(t, \cdot, \mu)$. To obtain the necessary a priori bounds on $\|\partial_{\mu x} V\|_{L^\infty}$ we rely on several representation formulas via suitable FBSDE systems. Although these representation formulas are similar in spirit to the ones used in [GMMZ22, MZ22a], we need to work with different systems of FBSDEs. During this process, we show also that – similarly as in [GMMZ22, MZ22a] – the small time horizon depends on the W_2-Lipschitz constant of $\partial_x V$ (and not on the W_1-Lipschitz constant as in [CD18b]).

This approach represents a major difference with the work [GMMZ22] and let us elaborate more on this. Indeed, we can observe that the FBSDE systems in [GMMZ22] (see for instance the [GMMZ22, System (2.24)]) are not natural if the intensity of the idiosyncratic noise is taken to be zero. Therefore, instead we will be working with FBSDE systems of Pontryagin type, where the natural variables are the state and the momentum (instead of the state and optimal value, as in [GMMZ22]). This system will become the classical forward-backward Hamiltonian system in case of deterministic problems. This subtlety has already been emphasized in [CCD22, Section 5.2]: in the case when non-degenerate idiosyncratic noise is present optimal paths may be characterized by solutions of FBSDEs, where the value function is represented as the decoupling field of the forward-backward system (similarly to the approach used also in [GMMZ22]); on the contrary, when the idiosyncratic noise is degenerate but additional convexity is present on the data (which is provided in our case by the displacement monotonicity), the natural characterization of the optimal paths may be obtained via the stochastic Pontryagin principle, where the decoupling field of the FBSDE system is understood as the gradient of the value function of the optimization problem.

The FBSDE systems used in [GMMZ22] required slightly stricter assumptions on the data. For instance, as we can see in [GMMZ22, Assumpotions 3.1 and 3.2], G was assumed to be globally Lipschitz continuous (with respect to the metric W_1 in the measure variable) and H was assumed to be Lipschitz continuous in all three variables (locally in the momentum variable, but globally in the state and measure variables). These actually imposed that $\partial_x G, \partial_\mu G$ are uniformly bounded (in $\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$) and in case of H, $\partial_x H, \partial_p H$ are uniformly bounded in $\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times B_R(0)$ and $\partial_\mu H$ uniformly bounded in $\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \times B_R(0)$ (with constants possibly depending on $R > 0$). In contrast to these, we improve these assumptions in the way that we require only $\partial_x G, \partial_x H, \partial_p H$ to be uniformly Lipschitz continuous (see Assumptions 4 and 5 below). As a result of the assumption in [GMMZ22] the value function was such that both $\partial_x V$ and $\partial_\mu V$ are uniformly bounded, while this will not be the case in our work (allowing for instance both $\partial_x V$ and $\partial_\mu V$ to have linear growth in x at infinity).

It is not hard to see that adding an additional idiosyncratic noise (with a constant intensity) would result in essentially the same analysis as the one present in this chapter. The Pontryagin principle used here is very similar in spirit to the analysis on the deterministic Hamiltonian system in [GM22a], however, the results there (as they rely both on the separable Hamiltonian and potential game structure) cannot imply our results if $\beta = 0$. Therefore, the results of our chapter unify and generalize the results of both [GM22a] and [GMMZ22].

The structure of the rest of the chapter is simple. In Section 3.2 we present some notations and the necessary assumptions on the data. Section 3.3 contains the classical semi-concavity estimates and convexity results for the master function. Here we discuss the propagation of the displacement monotonicity and its consequences as well. Section 3.4 contains the technical results on the representation formula for $\partial_{\mu x} V$ which yields the crucial a priori W_1-Lipschitz estimate for $\partial_x V$. Section 3.5 contains a by now standard argument describing how to extend the local in time well-posedness theory for the master equation, in case of sufficient a priori estimates.

3.2 Notations, setup and assumptions

We assume that we have a fixed standard filtered probability space $(\Omega, \mathbb{P}, \mathcal{F}_s)$. We use $L^2(\Omega)$ to denote the L^2 functions that map Ω into $\mathbb{R}^d$. We let $(B_t^0)_{t\in[0,T]}$ be a standard Brownian motion on $\mathbb{R}^d$ adapted to the filtration and define $B_t^{0,t_0} := B_t^0 - B_{t_0}^0$ which is a Brownian motion starting at the time t_0. We also let $\mathbb{L}^2(\mathcal{F}_s)$ be the collection of random variables with finite second moment that are measurable with respect to $\mathcal{F}_t$. We also use the notation $\mathbb{F}^0 := \{\mathcal{F}_s\}_{s\in[0,T]}$.

3.2.1 Standing assumptions

We will always assume that G, H are displacement monotone in the sense of Definition 2.2.1 and Definition 2.2.4, respectively. We also make the following regularity assumptions.

Assumption 4. We assume that

1. $G, \partial_x G, \partial_{xx} G \in \mathcal{C}^2(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d))$ and $\partial_\mu G, \partial_{x\mu} G \in \mathcal{C}^2(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d)$.
2. $|\partial_{xx} G|$ and $|\partial_{\mu x} G|$ are uniformly bounded by L^G.

We underline that these assumptions are weaker than those in [GMMZ22, Assumption 3.1] in that we do not require the uniform boundedness of $\partial_x G$ or $\partial_\mu G$.

We denote with L_2^G the Lipschitz constant of $\partial_x G$ with respect to space and W_2. Note the boundedness of $\partial_{\mu x} G$ implies that $\partial_x G$ is Lipschitz with respect to W_1 and so $L_2^G \leq L^G$.

Assumption 5. We assume that

1. $H \in \mathcal{C}^3(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d)$, $\partial_x H, \partial_p H, \partial_{xx} H, \partial_{xp} H, \partial_{pp} H, \partial_{xxp} H, \partial_{xpp} H, \partial_{ppp} H \in \mathcal{C}^2(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d)$ and $\partial_\mu H, \partial_{x\mu} H, \partial_{p\mu} H, \partial_{xp\mu} H, \partial_{pp\mu} H \in \mathcal{C}^2(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \times \mathbb{R}^d)$.
2. $\partial_{pp} H \geq c_0 I$ for some $c_0 > 0$.
3. $|\partial_{px} H|, |\partial_{xx} H|, |\partial_{pp} H|, |\partial_{x\mu} H|, |\partial_{p\mu} H|$ are uniformly bounded by L^H.

We remark that (2) in the above tells us that H is convex in the p variable.

Definition 3.2.1. A constant C is said to be universal if it depends only on the above quantities (L^G, L^H, and c_0) and T.

We also require some assumptions for the short time well-posedness of the master equation imposed in [CD18b, Theorem 5.45] (in particular those that are imposed in A2 of Assumption (MFG Smooth Coefficients) on page 414 which is required for the quoted theorem). We note that none of the a priori estimates depend on these.

Assumption 6. Let $w = (x, p)$. A derivative of H having linear growth will mean that is bounded by a constant times $1 + |w| + M_1(\mu)$. For a derivative of G linear growth will mean bounded by a constant times $1 + |x| + M_1(\mu)$.

For H we assume that

1. $\partial_{\mu\mu} H, \partial_{\mu} H, \partial_{\tilde{x}\mu} H$ exists and have linear growth.
2. $\partial_{www} H, \partial_{\mu\mu w} H, \partial_{\tilde{x}\mu w} H, \partial_{ww\mu} H$ exist and are bounded and Lipschitz continuous with respect to the space and measure variables (the latter with respect to W_1).

For G we assume that

1. $\partial_{\mu\mu} G, \partial_{\mu} G, \partial_{\tilde{x}\mu} G$ exist and have linear growth.
2. $\partial_{xxx} G, \partial_{\mu\mu x} G, \partial_{\tilde{x}\mu x} G, \partial_{xx\mu} G$ exist and are bounded and Lipschitz continuous with respect to the space and measure variables (the latter with respect to W_1).

3.2.2 The roadmap of the well-posedness theory

The proof of our main theorem (Theorem 3.5.3) will go as follows. First we assume that we have a smooth solution, V, to the master equation (3.1.2). We show that V is displacement monotone. This is a consequence of the displacement monotonicity assumption on H and G and will follow as the corresponding propagation of monotonicity result in [GMMZ22].

The heart of our analysis is to show that $\partial_x V$ is Lipschitz continuous (with respect to the Euclidean norm in x and the W_1 distance in μ) where the Lipschitz constant is universal. To achieve this, we will proceed as follows. First, the uniform boundedness of $\partial_{xx} V$ will be implied by semi-concavity estimates on V in the space variable (which comes from classical optimal control arguments) and its convexity in the space variable (which is implied immediately by displacement monotonicity, c.f. Lemma 2.2.3). Second, for the Lipschitz continuity of $\partial_x V$ in the measure variable we will first show that this is Lipschitz continuous with respect to the W_2 distance, which will also follow exactly as in [GMMZ22]. From this, together with the uniform bounds on $\partial_{xx}V$, we will be able to see that a certain system of FBSDEs is well-posed and that this system gives us a representation formula for $\partial_{x\mu}V$. From here we deduce that $\partial_{x\mu}V$ is bounded by a universal constant which is equivalent to $\partial_x V$ being W_1-Lipschitz continuous (with the same universal constant). The arguments to show that $\partial_{x\mu}V$ is uniformly bounded are different from the arguments in [GMMZ22], as they rely on the study of different linearized systems (than in [GMMZ22]) derived from the Pontryagin system.

Once these a priori bounds are proven, the well-posedness of the master equation will follow easily. Indeed by [CD18b] we have that the master equation is well-posed for short time and this short time depends only on the Lipschitz constant of $\partial_x G$ (with respect to space and W_1 norm in measure).

3.3 Semi-concavity and displacement monotonicity of the master function

3.3.1 Semi-concavity of value functions

Lemma 3.3.1. *Let V be a classical solution to the master equation. Fix some $t \in [0,T]$ and $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. Then there exists a path $(\rho_s)_{s\in[t,T]}$ of random probability measures with $\rho_t = \mu$ (specifically the ones given by the second equation of the mean field games system (3.1.2)) so that*

$$V(t,x,\mu) = \inf_{\alpha_s} \mathbb{E}\left\{ G(X_T,\rho_T) + \int_t^T L(X_s,\rho_s,\alpha_s(X_s))ds \right\},$$

where the infimum is taken over all pairs (X_s,α_s) that satisfy the SDE

$$dX_s = \alpha_s(X_s)ds + \beta dB_s^{0,t}, \text{ with } X_t = x.$$

Here $\mathbb{E}^x$ stands for the conditional expectation with respect to the event $X_t = x$.

Proof. This is essentially a folklore result discussed in [GMMZ22, Remark 2.10 part (ii)]. □

Proposition 3.3.2. *Let $L(x, s, v)$ be a stochastic process adapted to a filtration generated by a Brownian motion $(B_s^{0,t})_{s\in[t,T]}$ and $G(x)$ be a random variable that is measurable with respect to the filtration at T (in particular both L, G depend on ω which is suppressed in the notation).*

Suppose that L, G are semi-concave (in x) with a constant $C > 0$ and the optimal control problem

$$\inf_{\alpha} \mathbb{E}\left\{G(X_T) + \int_t^T L(X_s, s, \alpha_s)ds\right\}$$

$$\text{subject to } \begin{cases} dX_s = \alpha_s ds + \beta dB_s^{0,t}, & s \in (t, T) \\ X_t = x \end{cases}$$

has a solution (i.e. an optimal control $(\alpha_s)_{s\in[t,T]}$ which is an adapted process) for every x, t. Let

$$V(t, x) = \min_{\alpha} \mathbb{E}\left\{G(X_T) + \int_t^T L(X_s, s, \alpha_s)ds\right\}.$$

Then V is semi-concave with a semi-concavity constant $(1 + T)C$.

The results of this proposition are certainly well-known for experts (see for instance [CS04] for the deterministic setting, i.e. when $\beta = 0$, and [BCQ10] for a similar stochastic control problem). However we were unable to find a reference that matches our exact assumptions. Hence for completeness we reprove it.

Proof. Fix some (t, x) and let $(\alpha_s)_{s\in[t,T]}$ be the associated optimal control. Fix some $\lambda \in \mathbb{R}^d$. Consider the exact same control $(\alpha_s)_{s\in[t,T]}$ as a proposed control for the problem initiated at $(t, x + \lambda)$. Note that the solution to

$$\begin{cases} dX_s^1 = \alpha_s ds + \beta dB_s^{0,t}, & s \in (t, T), \\ X_t^1 = x + \lambda \end{cases}$$

is simply $X^1 = X + \lambda$.

We get

$$\begin{aligned} V(t, x+\lambda) &\leq \mathbb{E}\left\{G(X_T^1) + \int_t^T L(X_s^1, s, \alpha_s)ds\right\} \\ &= \mathbb{E}\left\{G(X_T + \lambda) + \int_t^T L(X_s + \lambda, s, \alpha_s)ds\right\} \end{aligned}$$

By a symmetric argument we get

$$V(t, x-\lambda) \leq \mathbb{E}\left\{G(X_T - \lambda) + \int_t^T L(X_s - \lambda, s, \alpha_s)ds\right\}$$

and so

$$\begin{aligned} &\frac{V(t, x+\lambda) + V(t, x-\lambda)}{2} - V(t, x) \\ &\leq \mathbb{E}\left\{\frac{G(X_T + \lambda) + G(X_T - \lambda)}{2} - G(X_T)\right\} \\ &+ \mathbb{E}\left\{\int_t^T \frac{L(X_s + \lambda, s, \alpha_s) + L(X_s - \lambda, s, \alpha_s)}{2} - L(X_s, s, \alpha_s)ds\right\} \\ &\leq C\left|\lambda\right|^2 + TC\left|\lambda\right|^2 \end{aligned}$$

as desired. □

Corollary 3.3.3. *If V is a classical solution to the master equation then V is semi-concave with a universal semi-concavity constant.*

Proof. By (3) in Assumption 5 (specifically the bound on $\partial_{xx}H$) we have that L is semi-concave (in x) with universal constant. □

3.3.2 Propagation of displacement monotonicity and a priori W_2-Lipschitz continuity

Proposition 3.3.4. *Suppose that G, H are displacement monotone and satisfy Assumptions 4 and 5 and that V is a classical solution to the master equation. Furthermore, assume* *that* $V(t,\cdot,\cdot)$, $\partial_x V(t,\cdot,\cdot)$, $\partial_{xx}V(t,\cdot,\cdot) \in \mathcal{C}^2(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d))$ *and* $\partial_\mu V(t,\cdot,\cdot,\cdot), \partial_{x\mu}V(t,\cdot,\cdot,\cdot) \in \mathcal{C}^2(\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d)$ *. Then for each fixed* t*, we have* $V(t,\cdot)$ *is displacement monotone. Furthermore* $\partial_x V$ *is Lipschitz continuous in* μ *with respect to the* W_2 *metric and the Lipschitz constant is universal.*

Remark 3.3.5. The additional regularity of V needed for this Proposition (specifically that $\partial_x V, \partial_{xx}V, \partial_\mu V, \partial_{x\mu}V \in \mathcal{C}^2$) will be provided below by Lemma 3.5.2.

Proof of Proposition 3.3.4. The claims follow the exact same proof as in [GMMZ22, Theorem 4.1, Theorem 5.1]. The reason that we do not need the bound on $\|\partial_x V\|_{L^\infty}$ is that in [GMMZ22, Theorem 5.1] this is only used via second derivatives of H, e.g. to get that $\partial_{xp}H(\cdot,\cdot,\partial_x V)$ is bounded. Under the assumptions in [GMMZ22, Theorem 5.1], $\partial_{xp}H$ is only locally bounded in p whereas we assume a uniform bound. Furthermore, [GMMZ22, Theorem 5.1] proves two results, that both $\partial_x V$ and V are W_2 Lipschitz whereas we only need that $\partial_x V$ is.

□

Corollary 3.3.6. *Suppose that* V *is a classical solution to the master equation. Then* $\partial_x V$ *is uniformly Lipschitz in space and measure variables with respect to the* W_2 *metric in the case of the measure component. Furthermore, the Lipschitz constant is universal.*

Proof. From Lemma 2.2.3 we have that $\partial_{xx}V \geq 0$ and from Corollary 3.3.3 we get that $\partial_{xx}V \leq CI$. Hence $|\partial_{xx}V|$ is bounded by a universal constant and so $\partial_x V$ is Lipschitz continuous in space with universal Lipschitz constant.

The Lipschitz continuity in measure comes from Proposition 3.3.4. □

3.4 A Priori W_1-Lipschitz estimates on $\partial_x V(x,\cdot)$

Several FBSDE systems will play a crucial role in our analysis.

3.4.1 FBSDE of Pontryagin type

Let $t_0 \in [0,T)$ and $\xi \in \mathbb{L}^2(\mathcal{F}_{t_0})$ and on the time interval $[t_0,T]$ we consider

$$\begin{cases} X_t^\xi &= \xi + \int_{t_0}^t \partial_p H(X_s^\xi, \rho_s, Y_s^\xi) ds + \beta B_t^{0,t_0} \\ Y_t^\xi &= \partial_x G(X_T^\xi, \rho_T) + \int_t^T \partial_x H(X_s^\xi, \rho_s, Y_s^\xi) ds + \int_t^T Z_s^{0,\xi} dB_s^{0,t_0} \end{cases} \tag{3.4.1}$$

where $\rho_{t_0} := \mathcal{L}_\xi$ and $\rho_t := \mathcal{L}_{X_t^\xi | \mathcal{F}_t^0}$.

Lemma 3.4.1. *Suppose that V is a classical solution to the master equation. Then we have the representation formulas $Y_t^\xi = \partial_x V(t, X_t^\xi, \rho_t)$ and*

$$Z_t^{0,\xi} = \beta \left(\partial_{xx} V(t, X_t^\xi, \rho_t) + \tilde{\mathbb{E}}_{\mathcal{F}_t}[\partial_{x\mu} V(t, X_t^\xi, \rho_t, \tilde{X}_t^\xi)] \right).$$

Proof. This result is well-known for experts and its proof follows the same lines as the proofs of [CD18b, Proposition 5.42], [CCD22, Remark 57], [GMMZ22, Theorem 6.3] and [MM24, Theorem 4.1]. □

We also consider the standard system

$$\begin{cases} X_t^x &= x + \beta B_t^{0,t_0} \\ Y_t^{x,\xi} &= \partial_x G(X_T^x, \rho_T) + \int_t^T \partial_x H(X_s^x, \rho_s, Y_s^{x,\xi}) ds + \int_t^T Z_s^{0,x,\xi} dB_s^{0,t_0} \end{cases} \tag{3.4.2}$$

and the alternative system

$$\begin{cases} X_t^{\xi,x} &= x + \int_{t_0}^t \partial_p H(X_s^{\xi,x}, \rho_s, Y_s^{\xi,x}) ds + \beta B_t^{0,t_0} \\ Y_t^{\xi,x} &= \partial_x G(X_T^{\xi,x}, \rho_T) + \int_t^T \partial_x H(X_s^{\xi,x}, \rho_s, Y_s^{\xi,x}) ds + \int_t^T Z_s^{0,\xi,x} dB_s^{0,t_0}, \end{cases} \tag{3.4.3}$$

which also have the corresponding representation formulas. Note the the difference between the variables $(Y_t^{x,\xi}, Z_s^{0,x,\xi})$ and $(Y_t^{\xi,x}, Z_s^{0,\xi,x})$ which is expressed in the superscript labels. We underline that the solutions of (3.4.2) and (3.4.3) depend implicitly on ξ via the flow of measures $(\rho_t)_{t\in[t_0,T]}$.

We emphasize that these differ from the systems in [GMMZ22] in that for us the variable Y plays the role of the momentum along the characteristics whereas in [GMMZ22] it is the value function along the characteristics. All the previous FBSDE systems presented above are strongly linked to the Pontryagin system associated to the stochastic maximum principle (or to the classical Hamiltonian system when $\beta = 0$), and the systems used in [GMMZ22] have no such connection.

Both previously presented systems have similar representation formulas as in Lemma 3.4.1. Indeed,

$$Y_t^{x,\xi} = \partial_x V(t, X_t^x, \rho_t), \quad Y_t^{\xi,x} = \partial_x V(t, X_t^{\xi,x}, \rho_t).$$

These are clearly different quantities, as in particular, if $\beta = 0$, we simply have $Y_t^{x,\xi} = \partial_x V(t, x, \rho_t) \neq \partial_x V(t, X_t^{\xi,x}, \rho_t)$, except when $t = t_0$, when $Y_{t_0}^{x,\xi} = Y_{t_0}^{\xi,x} = \partial_x V(t_0, x, \rho_{t_0})$.

3.4.2 Intuition

To help to give some intuition for the role of the different systems considered above, let us consider $\beta = 0$ for this subsection. In the deterministic case we can use $X_t^{\xi,x}$ to define ρ_t. In particular we get that $\rho_t = \mathcal{L}(X_t^{\xi,\xi(\cdot)}) = \mathcal{L}(X_t^{\xi,\cdot} \circ \xi)$ (note that $X_{t_0}^{\xi,x} = x$).

Our objective is to develop some equations that give a representation for $\partial_{\mu x} V$. Since $Y_t^{x,\xi} = \partial_x V(t, x, \rho_t)$ (since $\beta = 0$) it would be natural to try to differentiate the defining equation of $Y_t^{x,\xi}$ with respect to ξ (this will become (3.4.6) below). Let us formally attempt this. Our equation is

$$Y_t^{x,\xi} = \partial_x G(X_T^x, \rho_T) + \int_t^T \partial_x H(X_s^x, \rho_s, Y_s^{x,\xi}) ds$$

and so we see that this comes down to differentiating the flow of measures $\mathcal{L}(X_t^{\xi,\cdot} \circ \xi)$ with respect to ξ. Replacing ξ with $\xi + \epsilon e_1$ we get

$$\begin{aligned} X_s^{\xi+\epsilon e_1,\cdot}((\xi(\omega) + \epsilon e_1)) &\approx X_s^{\xi+\epsilon e_1,\cdot}(\xi(\omega)) + \epsilon \nabla X_s^{\xi,\cdot}(\xi(\omega)) \cdot e_1 \\ &\approx X_s^{\xi,\cdot}(\xi(\omega)) + \epsilon \nabla X_s^{\xi,\cdot}(\xi(\omega)) \cdot e_1 + \epsilon \delta X_s^{\xi,\cdot}(\xi(\omega)) \end{aligned}$$

where $\delta X_s^{\xi,\cdot}$ represents the value that one gets when from $X_s^{\xi,\cdot}$ when perturbing ξ. So we see that an equation that gives the variation of $Y_t^{x,\xi}$ with respect to ξ will involve two types of variations of X. The first is a gradient in space and the second is a variation with respect to ξ. Each of these will require their own system of FBSDEs which gives us three systems in total.

This also helps us understand the reason that we need to consider the three systems above (3.4.1),(3.4.2),(3.4.3). Having (3.4.2) is a matter of convenience as it provides the simplest representation formula. (3.4.3) is necessary because we need to understand the gradient in space of X. In the case of no noise these two alone would have been sufficient. However in the presence of noise we must also consider (3.4.1) because we cannot extract the ρ_s directly from (3.4.3).

3.4.3 FBSDEs for pointwise representation

In order to gain the necessary a priori regularity estimates on $\partial_x V$ (notably the fact that it is W_1–Lipschitz continuous in the measure variable), we work at the level of linearized FBSDE systems. These are derived from (3.4.1), (3.4.2) and (3.4.3). Linearization techniques combined with finite dimensional projections (in the measure variable) are underneath essentially all well-posedness results on master equations. This is typically carried out either at the PDE level using the MFG system (as for instance in [CDLL19, AM23], etc.) or at the level of the Hamiltonian/FBSDE system (as for instance in [CD18b, GS15, GM22a, GMMZ22], etc.)

Consider $\{e_1, \dots, e_d\} \subset \mathbb{R}^d$ the canonical basis and for $k \in \{1, \dots, d\}$. First, we differentiate (3.4.3) in the e_k direction to obtain

$$\begin{cases} \nabla_k X_t^{\xi,x} &= e_k + \int_{t_0}^t \left\{ (\nabla_k X_s^{\xi,x})^\top \partial_{xp} H(X_s^{\xi,x}, \rho_s, Y_s^{\xi,x}) + (\nabla_k Y_s^{\xi,x})^\top \partial_{pp} H(X_s^{\xi,x}, \rho_s, Y_s^{\xi,x}) \right\} ds \\ \nabla_k Y_t^{\xi,x} &= \partial_{xx} G(X_T^{\xi,x}, \rho_T) \cdot \nabla_k X_T^{\xi,x} \\ & \quad + \int_t^T \left\{ \partial_{xx} H(X_s^{\xi,x}, \rho_s, Y_s^{\xi,x}) \cdot \nabla_k X_s^{\xi,x} + \partial_{px} H(X_s^{\xi,x}, \rho_s, Y_s^{\xi,x}) \cdot \nabla_k Y_s^{\xi,x} \right\} ds \\ & \quad + \int_t^T \nabla_k Z_s^{0,\xi,x} \cdot dB_s^{0,t_0}. \end{cases} \tag{3.4.4}$$

$$\begin{cases} \nabla_k \mathcal{X}_t^{\xi,x} &= + \int_{t_0}^t \Big\{ (\nabla_k \mathcal{X}_s^{\xi,x})^\top \partial_{xp} H(X_s^{\xi}, \rho_s, Y_s^{\xi}) + (\nabla_k \mathcal{Y}_s^{\xi,x})^\top \partial_{pp} H(X_s^{\xi}, \rho_s, Y_s^{\xi}) \\ & \quad + \tilde{\mathbb{E}}_{\mathcal{F}_s} \big[(\nabla_k \tilde{X}_s^{\xi,x})^\top (\partial_{\mu p} H)(X_s^{\xi}, \rho_s, \tilde{X}_s^{\xi,x}, Y_s^{\xi}) \\ & \quad + (\nabla_k \tilde{\mathcal{X}}_s^{\xi,x})^\top \partial_{\mu p} H(X_s^{\xi}, \rho_s, \tilde{X}_s^{\xi}, Y_s^{\xi}) \big] \Big\} ds \\ \nabla_k \mathcal{Y}_t^{\xi,x} &= \partial_{xx} G(X_T^{\xi}, \rho_T) \cdot \nabla_k \mathcal{X}_T^{\xi,x} \\ & \quad + \tilde{\mathbb{E}}_{\mathcal{F}_T} \big[\partial_{\mu x} G(X_T^{\xi}, \rho_T, \tilde{X}_T^{\xi,x}) \cdot \nabla_k \tilde{X}_T^{\xi,x} + \partial_{\mu x} G(X_T^{\xi}, \rho_T, \tilde{X}_T^{\xi}) \cdot \nabla_k \tilde{\mathcal{X}}_T^{\xi,x} \big] \\ & \quad + \int_t^T \Big\{ \partial_{xx} H\big(X_s^{\xi}, \rho_s, Y_s^{\xi}\big) \cdot \nabla_k \mathcal{X}_s^{\xi,x} + \partial_{px} H\big(X_s^{\xi}, \rho_s, Y_s^{\xi}\big) \cdot \nabla_k \mathcal{Y}_s^{\xi,x} \\ & \quad + \tilde{\mathbb{E}}_{\mathcal{F}_s} \big[\partial_{\mu x} H\big(X_s^{\xi}, \rho_s, \tilde{X}_s^{\xi,x}, Y_s^{\xi}\big) \cdot \nabla_k \tilde{X}_s^{\xi,x} + \partial_{\mu x} H\big(X_s^{\xi}, \rho_s, \tilde{X}_s^{\xi}, Y_s^{\xi}\big) \cdot \nabla_k \tilde{\mathcal{X}}_s^{\xi,x} \big] \Big\} ds \\ & \quad + \int_t^T \nabla_k \mathcal{Z}_s^{0,\xi,x} \cdot dB_s^{0,t_0}, \end{cases} \tag{3.4.5}$$

and

$$\begin{aligned} \nabla_{\mu_k} Y_t^{x,\xi,\tilde{x}} &= \tilde{\mathbb{E}}_{\mathcal{F}_T} \big[\partial_{\mu x} G(X_T^x, \rho_T, \tilde{X}_T^{\xi,\tilde{x}}) \cdot \nabla_k \tilde{X}_T^{\xi,\tilde{x}} + \partial_{\mu x} G(X_T^x, \rho_T, \tilde{X}_T^{\xi}) \cdot \nabla_k \tilde{\mathcal{X}}_T^{\xi,\tilde{x}} \big] \\ & \quad + \int_t^T \Big\{ \partial_{px} H(X_s^x, \rho_s, Y_s^{x,\xi}) \cdot \nabla_{\mu_k} Y_s^{x,\xi,\tilde{x}} \\ & \quad + \tilde{\mathbb{E}}_{\mathcal{F}_s} \big[\partial_{\mu x} H(X_s^x, \rho_s, \tilde{X}_s^{\xi,\tilde{x}}, Y_s^{x,\xi}) \cdot \nabla_k \tilde{X}_s^{\xi,\tilde{x}} + \partial_{\mu x} H(X_s^x, \rho_s, \tilde{X}_s^{\xi}, Y_s^{x,\xi}) \cdot \nabla_k \tilde{\mathcal{X}}_s^{\xi,\tilde{x}} \big] \Big\} ds \\ & \quad + \int_t^T \nabla_{\mu_k} Z_s^{0,x,\xi,\tilde{x}} \cdot dB_s^{0,t_0}. \end{aligned} \tag{3.4.6}$$

Remark 3.4.2. The motivation behind the notation ∇_{μ_k} in (3.4.6) is that this is linked to the k^{th} component for the Wasserstein gradient. Later we will also use the notation ∂_{μ_k} with a similar purpose, i.e. $\partial_{\mu_k} F := \partial_\mu F \cdot e_k$, for any F regular enough.

Lemma 3.4.3. *There is a constant $\delta > 0$ so that whenever $T - t_0 < \delta$ the systems* (3.4.1), (3.4.4)*, and* (3.4.5) *have a unique solution, where δ depends only on L^H and L_2^G (these are the bounds on the second derivatives of H and the bounds on Lipschitz constant of $\partial_x G$, with respect to space and W_2 in measure). Furthermore the solutions to these systems are bounded by controlled quantities, specifically there is a constant C depending only on T, L^H, and L_2^G so that if A_t is one of $\nabla_k X_t^{\xi,x}, \nabla_k Y_t^{\xi,x}, \nabla_k \mathcal{X}_t^{\xi,x}$, or $\nabla_k \mathcal{Y}_t^{\xi,x}$ then*

$$\mathbb{E}\left[\sup_{s\in[t_0,T]} |A_s|^2\right] \leq C$$

It is crucial in the above lemma that the constant δ depends only on the W_2-Lipschitz constant of $\partial_x G$ and not on the W_1-Lipschitz constant.

Proof. The proof is similar to [GMMZ22, Proposition 6.2(i)].

The described short time existence and uniqueness for (3.4.1) follows from [CD18b, Theorem 5.4].

For (3.4.4) we use [Zha17, Theorem 8.2.1]. In particular we note that because (3.4.4) is linear (in $\nabla_k X_T^{\xi,x}$ and $\nabla_k Y_T^{\xi,x}$) the short time interval only depends on the absolute value of the coefficients which only include $\partial_{xx} G$ and second derivatives of H. Finally from the cited theorem we see that $\mathbb{E}\left[\sup_{s\in[t_0,T]} |A_s|^2\right]$ is bounded for $A_s = \nabla_k X_t^{\xi,x}$ or $\nabla_k Y_t^{\xi,x}$.

Next we consider (3.4.5). Again this is a linear FBSDE. The proof follows very similiarly to [Zha17, Theorem 8.2.1] however a modification is required.

Consider the standard mapping F given by y_s maps to $\nabla_k \mathcal{Y}_s^{\xi,x}$ where $\nabla_k \mathcal{Y}_s^{\xi,x}$ is the solution to

$$\begin{cases} \nabla_k \mathcal{X}_t^{\xi,x} & = \int_{t_0}^t \Big\{ (\nabla_k \mathcal{X}_s^{\xi,x})^\top \partial_{xp} H(X_s^\xi, \rho_s, Y_s^\xi) + (y_s)^\top \partial_{pp} H(X_s^\xi, \rho_s, Y_s^\xi) \\ & + \tilde{\mathbb{E}}_{\mathcal{F}_s} \big[(\nabla_k \tilde{X}_s^{\xi,x})^\top (\partial_{\mu p} H)(X_s^\xi, \rho_s, \tilde{X}_s^{\xi,x}, Y_s^\xi) \\ & + (\nabla_k \tilde{\mathcal{X}}_s^{\xi,x})^\top \partial_{\mu p} H(X_s^\xi, \rho_s, \tilde{X}_s^\xi, Y_s^\xi) \big] \Big\} ds \\ \nabla_k \mathcal{Y}_t^{\xi,x} & = \partial_{xx} G(X_T^\xi, \rho_T) \cdot \nabla_k \mathcal{X}_T^{\xi,x} \\ & + \tilde{\mathbb{E}}_{\mathcal{F}_T} \big[\partial_{\mu x} G(X_T^\xi, \rho_T, \tilde{X}_T^{\xi,x}) \cdot \nabla_k \tilde{X}_T^{\xi,x} + \partial_{\mu x} G(X_T^\xi, \rho_T, \tilde{X}_T^\xi) \cdot \nabla_k \tilde{\mathcal{X}}_T^{\xi,x} \big] \\ & + \int_t^T \Big\{ \partial_{xx} H\big(X_s^\xi, \rho_s, Y_s^\xi\big) \cdot \nabla_k \mathcal{X}_s^{\xi,x} + \partial_{px} H\big(X_s^\xi, \rho_s, Y_s^\xi\big) \cdot y_s \\ & + \tilde{\mathbb{E}}_{\mathcal{F}_s} \big[\partial_{\mu x} H\big(X_s^\xi, \rho_s, \tilde{X}_s^{\xi,x}, Y_s^\xi\big) \cdot \nabla_k \tilde{X}_s^{\xi,x} + \partial_{\mu x} H\big(X_s^\xi, \rho_s, \tilde{X}_s^\xi, Y_s^\xi\big) \cdot \nabla_k \tilde{\mathcal{X}}_s^{\xi,x} \big] \Big\} ds \\ & + \int_t^T \nabla_k \mathcal{Z}_s^{0,\xi,x} \cdot dB_s^{0,t_0}, \end{cases}$$

We will show that F is a contraction mapping under the norm given by $\|y_s\|^2 = \sup_s \mathbb{E}(|y_s|^2)$ when T is sufficiently small (for now assume $T < 1$). Indeed fix some y^1, y^2 denote by $\Delta y := y^1 - y^2$ and $\nabla_k \mathcal{X}^i$ be the solutions to the above system with $y_s = y_s^i$. Let $\Delta X := \nabla_k \mathcal{X}^1 - \nabla_k \mathcal{X}^2$ and $\Delta Y := F(y^1) - F(y^2)$. Applying Grönwall's inequality to the first equation in the system we see that ΔX satisfies $\|\Delta X\| \leq CT \|\Delta y\|$ where C depends only on L^H. From the second equation we see that ΔY satisfies the system

$$\begin{cases} \Delta Y_t & = \partial_{xx} G(X_T^\xi, \rho_T) \cdot \Delta X_T \\ & + \tilde{\mathbb{E}}_{\mathcal{F}_T} \big[\partial_{\mu x} G(X_T^\xi, \rho_T, \tilde{X}_T^\xi) \cdot \tilde{\Delta X}_T \big] \\ & + \int_t^T \Big\{ \partial_{xx} H\big(X_s^\xi, \rho_s, Y_s^\xi\big) \cdot \Delta X_s^{\xi,x} + \partial_{px} H\big(X_s^\xi, \rho_s, Y_s^\xi\big) \cdot \Delta y_s \\ & + \tilde{\mathbb{E}}_{\mathcal{F}_s} \big[\partial_{\mu x} H\big(X_s^\xi, \rho_s, \tilde{X}_s^\xi, Y_s^\xi\big) \cdot \tilde{\Delta X}_s^{\xi,x} \big] \Big\} ds \\ & + \int_t^T \Delta \mathcal{Z}_s^{0,\xi,x} \cdot dB_s^{0,t_0}, \end{cases}$$

The only term that may seem concerning is $\tilde{\mathbb{E}}_{\mathcal{F}_T} \big[\partial_{\mu x} G(X_T^\xi, \rho_T, \tilde{X}_T^\xi) \cdot \tilde{\Delta X}_T \big]$ since we want to claim that the short time interval depends only on the W_2-Lipschitz constant of $\partial_x G$ and not on the W_1-Lipschitz constant of G. However, note that

$$\left|\tilde{\mathbb{E}}_{\mathcal{F}_T}\left[\partial_{\mu x}G(X_T^{\xi},\rho_T,\tilde{X}_T^{\xi})\cdot \tilde{\Delta X_T}\right]\right| \leq L_2^G\mathbb{E}(\Delta X_T^2)^{\frac{1}{2}}$$

which follows from (2.1). It now follows from standard BSDE estimates (see [Zha17, Theorem 4.2.1]) that $\|\Delta Y\| \leq C_1T\|\Delta y\| + C_2\|\Delta X\|$ where C_1, C_2 depend only on L^H and L_2^G. Combining this with our estimate $\|\Delta X\| \leq CT\|\Delta y\|$ we see that F is a contraction mapping as long as T is sufficiently small and so there exists a unique solution to the above system. In particular the small time interval and corresponding bound will depend only on L_2^G and not $\|\partial_{x\mu}G\|_{L^\infty}$.

□

Corollary 3.4.4. $\mathbb{E}\left[\left|\nabla_{\mu_k}Y_{t_0}^{x,\xi,\tilde{x}}\right|^2\right]$ *is bounded by a universal constant.*

Proof. We see that $\nabla_{\mu_k}Y_{t_0}^{x,\xi,\tilde{x}}$ is the solution to a linear BSDE with coefficients that are bounded by universal constants. □

3.4.4 Proof of the representation formula

The proof is broken into four steps. In the first step we develop a system of FBSDE that gives a representation for $\mathbb{E}\big[\partial_{\mu x}V(t_0,x,\mu,\xi)\eta\big]$ where $\eta \in \mathbb{L}^2(\mathcal{F}_{t_0},\mathbb{R})$ is arbitrary. In the next two steps we prove the representation formula for discrete and absolutely continuous measures respectively. Finally we prove it for general measures.

Proposition 3.4.5. *Suppose that G, H are displacement monotone and satisfy Assumptions 4 and 5 and that V is a classical solution to the master equation. Then*

$$\partial_{\mu_k x}V(t_0,x,\mu,\tilde{x}) = \nabla_{\mu_k}Y_{t_0}^{x,\xi,\tilde{x}} \tag{3.4.7}$$

Proof. Step 1. For any $\xi \in \mathbb{L}^2(\mathcal{F}_{t_0}, \mu)$ and any scalar random variable $\eta \in \mathbb{L}^2(\mathcal{F}_{t_0}, \mathbb{R})$, following standard arguments and by the stability property of the involved systems we have

$$\lim_{\varepsilon\to 0} \mathbb{E}\left[\sup_{t_0\le t\le T}\left|\frac{1}{\varepsilon}\left[X_t^{\xi+\varepsilon\eta e_1} - X_t^{\xi}\right] - \delta X_t^{\xi,\eta e_1}\right|^2\right] = 0, \tag{3.4.8}$$

where $\left(\delta X^{\xi,\eta e_1}, \delta Y^{\xi,\eta e_1}, \delta Z^{0,\xi,\eta e_1}\right)$ satisfies the linear McKean–Vlasov FBSDE

$$\begin{cases} \delta X_t^{\xi,\eta e_1} &= \eta e_1 + \displaystyle\int_{t_0}^t \Big\{ (\delta X_s^{\xi,\eta e_1})^\top \partial_{xp} H\big(X_s^\xi, \rho_s, Y_s^\xi\big) + (\delta Y_s^{\xi,\eta e_1})^\top \partial_{pp} H\big(X_s^\xi, \rho_s, Y_s^\xi\big) \\ &\quad + \tilde{\mathbb{E}}_{\mathcal{F}_s}\big[\partial_{\mu p} H(X_s^\xi, \rho_s, \tilde{X}_s^\xi, Y_s^\xi) \cdot \delta \tilde{X}_s^{\xi,\eta e_1}\big]\Big\} ds \\ \delta Y_t^{\xi,\eta e_1} &= \partial_{xx} G(X_T^\xi, \rho_T) \cdot \delta X_T^{\xi,\eta e_1} + \tilde{\mathbb{E}}_{\mathcal{F}_T}\big[\partial_{\mu x} G(X_T^\xi, \rho_T, \tilde{X}_T^\xi) \cdot \delta \tilde{X}_T^{\xi,\eta e_1}\big] \\ &\quad + \displaystyle\int_t^T \Big\{ \partial_{xx} H\big(X_s^\xi, \rho_s, Y_s^\xi\big) \cdot \delta X_s^{\xi,\eta e_1} + \partial_{px} H\big(X_s^\xi, \rho_s, Y_s^\xi\big) \cdot \delta Y_s^{\xi,\eta e_1} \\ &\quad + \tilde{\mathbb{E}}_{\mathcal{F}_s}\big[\partial_{\mu x} H(X_s^\xi, \rho_s, \tilde{X}_s^\xi, Y_s^\xi) \cdot \delta \tilde{X}_s^{\xi,\eta e_1}\big]\Big\} ds + \displaystyle\int_t^T \delta Z_s^{0,\xi,\eta e_1} \cdot dB_s^{0,t_0}. \end{cases} \tag{3.4.9}$$

Specifically let $\delta\Phi_t^{\xi,\eta e_1,\varepsilon} = \frac{1}{\varepsilon}(\Phi_t^{\xi+\varepsilon\eta e_1} - \Phi_t^\xi)$ for $\Phi \in \{X, Y, Z^0\}$. By substituting and subtracting in (3.4.1) we see

$$\begin{aligned} \delta X_t^{\xi,\eta e_1,\varepsilon} &= \eta e_1 + \frac{1}{\varepsilon}\int_{t_0}^t \Big\{\partial_p H(X_s^{\xi+\varepsilon\eta e_1}, \rho_s^{\xi+\varepsilon\eta e_1}, Y_s^{\xi+\varepsilon\eta e_1}) - \partial_p H(X_s^\xi, \rho_s^\xi, Y_s^\xi)\Big\} ds \\ &= \eta e_1 + \int_{t_0}^t \Big\{ (\delta X_s^{\xi,\eta e_1,\varepsilon})^\top \partial_{xp} H\big(X_s^\xi, \rho_s, Y_s^\xi\big) + (\delta Y_s^{\xi,\eta e_1,\varepsilon})^\top \partial_{pp} H\big(X_s^\xi, \rho_s, Y_s^\xi\big) \\ &\quad + \tilde{\mathbb{E}}_{\mathcal{F}_s}\big[\partial_{\mu p} H(X_s^\xi, \rho_s, \tilde{X}_s^\xi, Y_s^\xi) \cdot \delta \tilde{X}_s^{\xi,\eta e_1,\varepsilon}\big]\Big\} ds + O(\varepsilon) \end{aligned}$$

and

$$\begin{aligned} \delta Y_t^{\xi,\eta e_1,\varepsilon} &= \partial_{xx} G(X_T^\xi, \rho_T) \cdot \delta X_T^{\xi,\eta e_1,\varepsilon} + \tilde{\mathbb{E}}_{\mathcal{F}_T}\big[\partial_{\mu x} G(X_T^\xi, \rho_T, \tilde{X}_T^\xi) \cdot \delta \tilde{X}_T^{\xi,\eta e_1,\varepsilon}\big] \\ &\quad - \int_t^T \Big\{ \partial_{xx} H\big(X_s^\xi, \rho_s, Y_s^\xi\big) \cdot \delta X_s^{\xi,\eta e_1,\varepsilon} + \partial_{px} H\big(X_s^\xi, \rho_s, Y_s^\xi\big) \cdot \delta Y_s^{\xi,\eta e_1,\varepsilon} \\ &\quad + \tilde{\mathbb{E}}_{\mathcal{F}_s}\big[\partial_{\mu x} H(X_s^\xi, \rho_s, \tilde{X}_s^\xi, Y_s^\xi) \cdot \delta \tilde{X}_T^{\xi,\eta e_1,\varepsilon}\big]\Big\} ds + \int_t^T \delta Z_s^{0,\xi,\eta e_1,\varepsilon} \cdot dB_s^0 + O(\varepsilon) \end{aligned}$$

where the implicit constant in $O(\varepsilon)$ is bounded by various third derivatives of H (such as $\partial_{xxp} H, \partial_{xpp} H$, etc.) which are assumed to be bounded by Assumption 6. Note that aside from the $O(\varepsilon)$ term $\left(\delta X_t^{\xi,\eta e_1,\varepsilon}, \delta Y_t^{\xi,\eta e_1,\varepsilon}, \delta Z^{0,\xi,\eta e_1,\varepsilon}\right)$ satisfies the exact same FBSDE system as

$$\left(\delta X_t^{\xi,\eta e_1}, \delta Y_t^{\xi,\eta e_1}, \delta Z^{0,\xi,\eta e_1}\right).$$

By the stability of FBSDE we get

$$\lim_{\varepsilon\to 0} \mathbb{E}\left[\sup_{t_0\le t\le T}\left|\delta X_t^{\xi,\eta e_1,\varepsilon} - \delta X_t^{\xi,\eta e_1}\right|^2\right] = 0$$

as desired.

Similarly to (3.4.8), using (3.4.2), one can show that

$$\lim_{\varepsilon\to 0} \mathbb{E}\left[\sup_{t_0\le t\le T}\left|\frac{1}{\varepsilon}\left[Y_t^{x,\xi+\varepsilon\eta e_1} - Y_t^{x,\xi}\right] - \delta Y_t^{x,\xi,\eta e_1}\right|^2\right] = 0, \tag{3.4.10}$$

where $\left(\delta Y^{x,\xi,\eta e_1}, \delta Z^{0,x,\xi,\eta e_1}\right)$ satisfies the linear (standard) BSDE

$$\begin{aligned}\delta Y_t^{x,\xi,\eta e_1} &= \tilde{\mathbb{E}}_{\mathcal{F}_T^0}\left[\partial_{\mu x} G(X_T^x, \rho_T, \tilde{X}_T^{\xi})\cdot \delta\tilde{X}_T^{\xi,\eta e_1}\right] + \int_t^T \delta Z_s^{0,x,\xi,\eta e_1}\cdot dB_s^0 \\ + \int_t^T &\left\{\partial_{px} H(X_s^x, \rho_s, Y_s^{x,\xi})\cdot \delta Y_s^{x,\xi,\eta e_1} + \tilde{\mathbb{E}}_{\mathcal{F}_s}\left[\partial_{\mu x} H(X_s^x, \rho_s, \tilde{X}_s^{\xi}, Y_s^{x,\xi})\cdot \delta\tilde{X}_s^{\xi,\eta e_1}\right]\right\} ds\end{aligned} \tag{3.4.11}$$

In particular, (3.4.10) implies,

$$\lim_{\varepsilon\to 0}\left|\frac{1}{\varepsilon}\left[\partial_x V(t_0, x, \mathcal{L}_{\xi+\varepsilon\eta e_1}) - \partial_x V(t_0, x, \mathcal{L}_{\xi})\right] - \delta Y_{t_0}^{x,\xi,\eta e_1}\right|^2 = 0. \tag{3.4.12}$$

Thus, by the definition of $\partial_{\mu x} V$,

$$\mathbb{E}\left[\partial_{\mu x} V(t_0, x, \mu, \xi)\eta e_1\right] = \mathbb{E}\left[\partial_{\mu_1 x} V(t_0, x, \mu, \xi)\eta\right] = \delta Y_{t_0}^{x,\xi,\eta e_1}. \tag{3.4.13}$$

Step 2. In this step we assume that ξ (or say, μ) is discrete: $p_i = \mathbb{P}(\xi = x_i)$, $i = 1, \cdots, n$. In particular, we have that $\mu = \sum_{i=1}^n p_i x_i$, for some $\{x_1, \ldots, x_n\} \subset \mathbb{R}^d$. Fix i and consider the following system of McKean–Vlasov FBSDEs: for $j = 1, \cdots, n$,

$$
\begin{cases}
\nabla_{\mu_1} X_t^{i,j} &= \delta_{ij} e_1 + \displaystyle\int_{t_0}^t \Big\{ \sum_{k=1}^n p_k \tilde{\mathbb{E}}_{\mathcal{F}_s} \Big[(\nabla_{\mu_1} \tilde{X}_s^{i,k})^\top \partial_{\mu p} H(X_s^{\xi,x_j}, \rho_s, \tilde{X}_T^{\xi,x_k}, Y_s^{\xi,x_j}) \Big] \\
&\quad + (\nabla_{\mu_1} X_s^{i,j})^\top \partial_{xp} H\big(X_s^{\xi,x_j}, \rho_s, Y_s^{\xi,x_j}\big) + (\nabla_{\mu_1} Y_s^{i,j})^\top \partial_{pp} H\big(X_s^{\xi,x_j}, \rho_s, Y_s^{\xi,x_j}\big) \Big\} ds, \\
\nabla_{\mu_1} Y_t^{i,j} &= \partial_{xx} G(X_T^{\xi,x_j}, \rho_T) \cdot \nabla_{\mu_1} X_T^{i,j} + \displaystyle\sum_{k=1}^n p_k \tilde{\mathbb{E}}_{\mathcal{F}_T} \Big[\partial_{\mu x} G(X_T^{\xi,x_j}, \rho_T, \tilde{X}_T^{\xi,x_k}) \cdot \nabla_{\mu_1} \tilde{X}_T^{i,k} \Big] \\
&\quad + \displaystyle\int_t^T \Big\{ \partial_{xx} H\big(X_s^{\xi,x_j}, \rho_s, Y_s^{\xi,x_j}\big) \cdot \nabla_{\mu_1} X_s^{i,j} + \partial_{px} H\big(X_s^{\xi,x_j}, \rho_s, Y_s^{\xi,x_j}\big) \cdot \nabla_{\mu_1} Y_s^{i,j} \\
&\quad + \displaystyle\sum_{k=1}^n p_k \tilde{\mathbb{E}}_{\mathcal{F}_s} \Big[\partial_{\mu x} H(X_s^{\xi,x_j}, \rho_s, \tilde{X}_s^{\xi,x_k}, Y_s^{\xi,x_j}) \cdot \nabla_{\mu_1} \tilde{X}_s^{i,k} \Big\} ds \\
&\quad + \displaystyle\int_t^T \nabla_{\mu_1} Z_s^{0,i,j} \cdot dB_s^{0,t_0},
\end{cases}
\tag{3.4.14}
$$

where δ_{ij} stands for Kronecker's symbol. In the above system $\nabla_{\mu_1} X_t^{i,j}$ represents perturbing x_i in μ in the e_1 direction and measuring the variation in X_t at X^{ξ,x_j} (the place where x_j has moved to by time t). The interpretation for $\nabla_{\mu_1} Y_t^{i,j}$ is similar.

For any $\Phi \in \{X, Y, Z^0\}$, we define

$$
\nabla_1 \Phi^{\xi,x_i} := \nabla_{\mu_1} \Phi^{i,i}, \quad \nabla_1 \Phi^{\xi,x_i,*} := \frac{1}{p_i} \sum_{j \neq i} \nabla_{\mu_1} \Phi^{i,j} 1_{\{\xi = x_j\}}.
$$

Note that $\Phi^\xi = \sum_{j=1}^n \Phi^{\xi,x_j} 1_{\{\xi = x_j\}}$. Since (3.4.14) is linear, one can easily check that

$$
\begin{aligned}
\nabla_1 X_t^{\xi,x_i} = e_1 + \int_{t_0}^t &\Big\{ (\nabla_1 X_s^{\xi,x_i})^\top \partial_{xp} H\big(X_s^{\xi,x_i}, \rho_s, Y_s^{\xi,x_i}\big) + (\nabla_1 Y_s^{\xi,x_i})^\top \partial_{pp} H\big(X_s^{\xi,x_i}, \rho_s, Y_s^{\xi,x_i}\big) \\
&+ p_i \tilde{\mathbb{E}}_{\mathcal{F}_s} \Big[(\nabla_1 \tilde{X}_s^{\xi,x_i})^\top \partial_{\mu p} H(X_s^{\xi,x_i}, \rho_s, \tilde{X}_s^{\xi,x_i}, Y_s^{\xi,x_i}) \\
&+ (\nabla_1 \tilde{X}_s^{\xi,x_i,*})^\top \partial_{\mu p} H(X_s^{\xi,x_i}, \rho_s, \tilde{X}_s^{\xi}, Y_s^{\xi,x_i} \Big] \Big\} ds,
\end{aligned}
\tag{3.4.15}
$$

$$
\begin{aligned}
\nabla_1 X_t^{\xi,x_i,*} = -\int_{t_0}^t &\Big\{ (\nabla_1 X_s^{\xi,x_i,*})^\top \partial_{xp} H\big(X_s^{\xi}, \rho_s, Y_s^{\xi}\big) + (\nabla_1 Y_s^{\xi,x_i,*})^\top \partial_{pp} H\big(X_s^{\xi}, \rho_s, Y_s^{\xi}\big) \\
&+ \tilde{\mathbb{E}}_{\mathcal{F}_s} \Big[(\nabla_1 \tilde{X}_s^{\xi,x_i})^\top \partial_{\mu p} H(X_s^{\xi}, \rho_s, \tilde{X}_s^{\xi,x_i}, Y_s^{\xi}) \\
&+ (\nabla_1 \tilde{X}_s^{\xi,x_i,*})^\top \partial_{\mu p} H(X_s^{\xi}, \rho_s, \tilde{X}_s^{\xi}, Y_s^{\xi}) \Big] 1_{\{\xi \neq x_i\}} \Big\} ds
\end{aligned}
\tag{3.4.16}
$$

$$
\begin{aligned}
\nabla_1 Y_t^{\xi,x_i} &= \partial_{xx} G(X_T^{\xi,x_i}, \rho_T) \cdot \nabla_1 X_T^{\xi,x_i} - \int_t^T \nabla_1 Z_s^{0,\xi,x_i} dB_s^{0,t_0} \\
&+ p_i \tilde{\mathbb{E}}_{\mathcal{F}_T}\big[\partial_{\mu x} G(X_T^{\xi,x_i}, \rho_T, \tilde{X}_T^{\xi,x_i}) \cdot \nabla_1 \tilde{X}_s^{\xi,x_i} + \partial_{\mu x} G(X_T^{\xi,x_i}, \rho_T, \tilde{X}_T^{\xi}) \cdot \nabla_1 \tilde{X}_T^{\xi,x_i,*}\big] \\
&+ \int_t^T \Big\{\partial_{xx} H\big(X_s^{\xi,x_i}, \rho_s, Y_s^{\xi,x_i}\big) \cdot \nabla_1 X_s^{\xi,x_i} + \partial_{px} H\big(X_s^{\xi,x_i}, \rho_s, Y_s^{\xi,x_i}\big) \cdot \nabla_1 Y_s^{\xi,x_i} \\
&+ p_i \tilde{\mathbb{E}}_{\mathcal{F}_s}\big[\partial_{\mu x} H(X_s^{\xi,x_i}, \rho_s, \tilde{X}_s^{\xi,x_i}, Y_s^{\xi,x_i}) \cdot \nabla_1 \tilde{X}_s^{\xi,x_i} \\
&+ \partial_{\mu x} H(X_s^{\xi,x_i}, \rho_s, \tilde{X}_s^{\xi}, Y_s^{\xi,x_i}) \cdot \nabla_1 \tilde{X}_T^{\xi,x_i,*})\big]\Big\} ds
\end{aligned}
\tag{3.4.17}
$$

$$
\begin{aligned}
\nabla_1 Y_t^{\xi,x_i,*} &= \partial_{xx} G(X_T^{\xi}, \rho_T) \cdot \nabla_1 X_T^{\xi,x_i,*} - \int_t^T \nabla_1 Y_s^{0,\xi,x_i,*} \cdot dB_s^{0,t_0} \\
&+ \tilde{\mathbb{E}}_{\mathcal{F}_T}\big[\partial_{\mu x} G(X_T^{\xi}, \rho_T, \tilde{X}_T^{\xi,x_i}) \cdot \nabla_1 \tilde{X}_s^{\xi,x_i} + \partial_{\mu x} G(X_T^{\xi}, \rho_T, \tilde{X}_T^{\xi}) \cdot \nabla_1 \tilde{X}_T^{\xi,x_i,*}\big] 1_{\{\xi \neq x_i\}} \\
&+ \int_t^T \Big\{\partial_{xx} H\big(X_s^{\xi}, \rho_s, Y_s^{\xi}\big) \cdot \nabla_1 X_s^{\xi,x_i,*} + \partial_{px} H\big(X_s^{\xi}, \rho_s, Y_s^{\xi}\big) \cdot \nabla_1 Y_s^{\xi,x_i,*} \\
&+ \tilde{\mathbb{E}}_{\mathcal{F}_s}\big[\partial_{\mu x} H(X_s^{\xi}, \rho_s, \tilde{X}_s^{\xi,x_i}, Y_s^{\xi}) \cdot \nabla_1 \tilde{X}_s^{\xi,x_i} \\
&+ \partial_{\mu x} H(X_s^{\xi}, \rho_s, \tilde{X}_s^{\xi}, Y_s^{\xi}) \cdot \nabla_1 \tilde{X}_T^{\xi,x_i-})\big] 1_{\{\xi \neq x_i\}}\Big\} ds
\end{aligned}
\tag{3.4.18}
$$

Since (3.4.9) is also linear, one can check that, for $\Phi \in \{X, Y, Z^0\}$,

$$
\delta \Phi^{\xi, 1_{\{\xi = x_i\}} e_1} = \nabla_1 \Phi^{\xi,x_i} 1_{\{\xi = x_i\}} + p_i \nabla_1 \Phi^{\xi,x_i,*}. \tag{3.4.19}
$$

Moreover, note that

$$
\begin{aligned}
&\tilde{\mathbb{E}}_{\mathcal{F}_T}\big[\partial_{\mu x} G(X_T^{x,\xi}, \rho_T, \tilde{X}_T^{\xi}) \cdot \delta \tilde{X}_T^{\xi, 1_{\{\xi = x_i\}} e_1}\big] \\
&= \tilde{\mathbb{E}}_{\mathcal{F}_T}\Big[\partial_{\mu x} G(X_T^{x,\xi}, \rho_T, \tilde{X}_T^{\xi}) \cdot \big[\nabla_1 \tilde{X}_T^{\xi,x_i} 1_{\{\xi = x_i\}} + p_i \nabla_1 \tilde{X}_T^{\xi,x_i,*}\big]\Big] \\
&= p_i \tilde{\mathbb{E}}_{\mathcal{F}_T}\Big[\partial_{\mu x} G(X_T^{x,\xi}, \rho_T, \tilde{X}_T^{\xi,x_i}) \cdot \nabla_1 \tilde{X}_T^{\xi,x_i} + \partial_{\mu x} G(X_T^{x,\xi}, \rho_T, \tilde{X}_T^{\xi}) \cdot \nabla_1 \tilde{X}_T^{\xi,x_i,*}\Big]
\end{aligned}
$$

and similarly

$$
\begin{aligned}
&\tilde{\mathbb{E}}_{\mathcal{F}_s}\big[\partial_{\mu x} H(X_s^{x,\xi}, \rho_s, \tilde{X}_s^{\xi}, Y_s^{x,\xi}) \cdot \delta \tilde{X}_s^{\xi, 1_{\{\xi = x_i\}} e_1}\big] \\
&= p_i \tilde{\mathbb{E}}_{\mathcal{F}_s}\Big[\partial_{\mu x} H(X_s^{x,\xi}, \rho_s, \tilde{X}_s^{\xi,x_i}, Y_s^{x,\xi}) \cdot \nabla_1 \tilde{X}_s^{\xi,x_i} + \partial_{\mu x} H(X_s^{x,\xi}, \rho_s, \tilde{X}_s^{\xi}, Y_s^{x,\xi}) \cdot \nabla_1 \tilde{X}_s^{\xi,x_i,*})\Big].
\end{aligned}
$$

Plug this into (3.4.11), we obtain

$$\delta\Phi_t^{x,\xi,1_{\{\xi=x_i\}}e_1} = p_i \nabla_{\mu_1}\Phi_t^{x,\xi,x_i}, \tag{3.4.20}$$

where

$$\begin{aligned}\nabla_{\mu_1} Y_t^{x,\xi,x_i} &= \tilde{\mathbb{E}}_{\mathcal{F}_T}\big[\partial_{\mu x} G(X_T^x, \rho_T, \tilde{X}_T^{\xi,x_i}) \cdot \nabla_1 \tilde{X}_T^{\xi,x_i} + \partial_{\mu x} G(X_T^x, \rho_T, \tilde{X}_T^{\xi}) \cdot \nabla_1 \tilde{X}_T^{\xi,x_i,*})\big] \\ &+ \int_t^T \Big\{ \partial_{px} H(X_s^{x,\xi}, \rho_s, Y_s^{x,\xi}) \cdot \delta Y_s^{x,\xi,x_i} \\ &+ \tilde{\mathbb{E}}_{\mathcal{F}_s}\Big[\partial_{\mu x} H(X_s^{x,\xi}, \rho_s, \tilde{X}_s^{\xi,x_i}, Y_s^{x,\xi}) \cdot \nabla_1 \tilde{X}_s^{\xi,x_i} \\ &+ \partial_{\mu x} H(X_s^{x,\xi}, \rho_s, \tilde{X}_s^{\xi}, Y_s^{x,\xi}) \cdot \nabla_1 \tilde{X}_s^{\xi,x_i,*}\Big]\Big\} ds \\ &+ \int_t^T \nabla_{\mu_1} Z_s^{0,x,\xi,x_i} \cdot dB_s^{0,t_0}.\end{aligned} \tag{3.4.21}$$

In particular, by setting $\eta = 1_{\{\xi=x_i\}}$ in (3.4.13) we obtain:

$$\partial_{\mu_1 x} V(0, x, \mu, x_i) = \nabla_{\mu_1} Y_0^{x,\xi,x_i}. \tag{3.4.22}$$

We shall note that (3.4.15)-(3.4.16), (3.4.17)-(3.4.18) is different from (3.4.4) and (3.4.5), so (3.4.22) provides an alternative discrete representation.

Step 3. We now prove (3.4.7) in the case that μ is absolutely continuous. For each $n \geq 3$, set

$$x_{\vec{i}}^n := \frac{\vec{i}}{n}, \quad \Delta_{\vec{i}}^n := \left[\frac{i_1}{n}, \frac{i_1+1}{n}\right) \times \cdots \times \left[\frac{i_d}{n}, \frac{i_d+1}{n}\right), \quad \vec{i} = (i_1, \cdots, i_d)^\top \in \mathbb{Z}^d.$$

For any $x \in \mathbb{R}^d$, there exists $\vec{i}(x) := (i_1(x), \cdots, i_d(x)) \in \mathbb{Z}^d$ such that $x \in \Delta_{\vec{i}(x)}^n$. Let

$$\vec{i}^n(x) := (i_1^n(x), \cdots, i_d^n(x)) \in \mathbb{Z}^d, \quad \text{where} \quad i_l^n(x) := \min\{\max\{i_l, -n^2\}, n^2\},\ l = 1, \cdots, d.$$

Denote $Q_n := \{x \in \mathbb{R}^d : |x_i| \leq n, i = 1, \cdots, d\}$, $\mathbb{Z}_n^d := \{\vec{i} \in \mathbb{Z}^d \,:\, \Delta_{\vec{i}}^n \cap Q_n \neq \emptyset\}$, and

$$\xi_n := \sum_{\vec{i} \in \mathbb{Z}_n^d} x_{\vec{i}}^n 1_{\Delta_{\vec{i}}^n}(\xi) + \frac{\vec{i}^n(\xi)}{n} 1_{Q_n^c}(\xi). \tag{3.4.23}$$

It is clear that $\lim_{n\to+\infty}\mathbb{E}\big[|\xi_n-\xi|^2\big]=0$ and thus $\lim_{n\to\infty}W_2(\mathcal{L}_{\xi_n},\mathcal{L}_\xi)=0$. Then for any scalar random variable η, by stability of FBSDE (3.4.9) and BSDE (3.4.11), we derive from (3.4.13) that

$$\mathbb{E}\Big[\partial_{\mu_1 x}V(0,x,\mu,\xi)\eta\Big]=\delta Y_0^{x,\xi,\eta e_1}=\lim_{n\to\infty}\delta Y_0^{x,\xi_n,\eta e_1}. \tag{3.4.24}$$

For each $\tilde{x}\in\mathbb{R}^d$, let $\vec{i}(\tilde{x})$ be the i such that $\tilde{x}\in\Delta^n_{\vec{i}}$, which holds when $n>|\tilde{x}|$. Then $\big(\mathcal{L}_{\xi_n},\frac{\vec{i}(\tilde{x})}{n}\big)\to(\mu,\tilde{x})$ as $n\to\infty$ in W_2 and as a sequence in $\mathbb{R}^d$, respectively. By the stability of FBSDEs (3.4.1)-(3.4.2), we have as $n\to\infty$ that $X^{\xi_n,\frac{\vec{i}(\tilde{x})}{n}}\to X^{\xi,\tilde{x}}$ and $Y^{\xi_n,\frac{\vec{i}(\tilde{x})}{n}}\to Y^{\xi,\tilde{x}}$, as $n\to+\infty$, under the norm given by $\|A\|:=\mathbb{E}\left(\sup_t|A_t|^2\right)$. Moreover, since μ is absolutely continuous,

$$\mathbb{P}\Big(\xi_n=\frac{\vec{i}(\tilde{x})}{n}\Big)=\mathbb{P}\Big(\xi\in\Delta^n_{\vec{i}}\Big)\to 0,\quad \text{as}\quad n\to\infty.$$

Then by the stability of (3.4.15)-(3.4.16), (3.4.17)-(3.4.18) and (3.4.21) we can check that

$$\lim_{n\to\infty}\left(\nabla_1\Phi^{\xi_n,\frac{\vec{i}(\tilde{x})}{n}},\ \nabla_1\Phi^{\xi_n,\frac{\vec{i}(\tilde{x})}{n},*},\ \nabla_{\mu_1}\Phi^{x,\xi_n,\frac{\vec{i}(\tilde{x})}{n}}\right)=\left(\nabla_1\Phi^{\xi,\tilde{x}},\ \nabla_1\Phi^{\xi,\tilde{x},*},\ \nabla_{\mu_1}\Phi^{x,\xi,\tilde{x}}\right). \tag{3.4.25}$$

Now for any bounded function $\varphi\in C(\mathbb{R}^d)$, set $\eta=\varphi(\xi)$ in (3.4.24), we derive from (3.4.20) that

$$\mathbb{E}\Big[\partial_{\mu_1 x}V(0,x,\mu,\xi)\varphi(\xi)\Big]=\lim_{n\to\infty}\delta Y_0^{x,\xi_n,\varphi(\xi_n)e_1}=\lim_{n\to\infty}\sum_{\vec{i}\in\mathbb{Z}^d_n}\varphi\Big(x^n_{\vec{i}}\Big)\delta Y_0^{x,\xi_n,1_{\{\xi_n=x^n_{\vec{i}}\}}e_1}$$

and so,

$$\mathbb{E}\Big[\partial_{\mu_1 x}V(0,x,\mu,\xi)\varphi(\xi)\Big]=\lim_{n\to\infty}\sum_{\vec{i}\in\mathbb{Z}^d_n}\varphi(x^n_{\vec{i}})\nabla_{\mu_1}Y_0^{x,\xi_n,x^n_{\vec{i}}}\mathbb{P}(\xi\in\Delta_{\vec{i}})=\int_{\mathbb{R}^d}\varphi(\tilde{x})\nabla_{\mu_1}Y_0^{x,\xi,\tilde{x}}d\mu(\tilde{x}).$$

This implies (3.4.7) immediately.

Step 4. We finally prove the general case. Denote $\psi(x,\mu,\tilde{x}):=\nabla_{\mu_1}Y_0^{x,\xi,\tilde{x}}$. By the stability of FBSDEs, ψ is continuous in all the variables. Fix an arbitrary (μ,ξ). One can construct ξ_n such that $\mathcal{L}_{\xi_n}$ is absolutely continuous and $\lim_{n\to\infty}\mathbb{E}[|\xi_n-\xi|^2]=0$. Then, for any $\eta=\varphi(\xi)$ as in Step 3, by (3.4.13) and Step 3 we have

$$\begin{aligned}&\mathbb{E}\big[\partial_{\mu_1 x}V(0,x,\mu,\xi)\varphi(\xi)\big]=\lim_{n\to\infty}\delta Y_0^{x,\xi_n,\varphi(\xi_n)e_1}\\&=\lim_{n\to\infty}\mathbb{E}\big[\psi(x,\mathcal{L}_{\xi_n},\xi_n)\varphi(\xi_n)\big]=\mathbb{E}\big[\psi(x,\mu,\xi)\varphi(\xi)\big],\end{aligned}$$

which implies (3.4.7) in the general case.

□

Corollary 3.4.6. *Suppose that G, H are displacement monotone and satisfy Assumptions 4 and 5 and that V is a classical solution to the master equation. Then $\partial_{x\mu} V$ is uniformly bounded by a universal constant.*

Proof. From the above representation formula we have $\partial_{\mu_k x} V(t_0, x, \mu, \tilde{x}) = \nabla_{\mu_k} Y_{t_0}^{x,\xi,\tilde{x}}$. In particular $\partial_{\mu_k x} V(t_0, x, \mu, \tilde{x})$ is deterministic and so

$$|\partial_{\mu_k x} V(t_0, x, \mu, \tilde{x})|^2 = \mathbb{E}(|\partial_{\mu_k x} V(t_0, x, \mu, \tilde{x})|^2) = \mathbb{E}\left(\left|\nabla_{\mu_k} Y_{t_0}^{x,\xi,\tilde{x}}\right|^2\right)$$

which is bounded by a universal constant by Corollary 3.4.4. □

3.5 Long Time Well-Posedness for the Master Equation

First we recall a short time existence result, [CD18b, Theorem 5.45].

Lemma 3.5.1. *Suppose that Assumptions 4, 5, and 6 are satisfied. Then there exists a universal constant $c > 0$ and V so that V is a classical solution to the master equation on $[T-c, T] \times \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$. Furthermore for each fixed $t \in [T-c, T]$, $V(t, \cdot, \cdot)$ satisfies the same assumptions as G in Assumption 6.*

Proof. First note that Assumption 6 gives the regularity conditions for the short time existence that will not affect the size of the time interval (in the notation of [CD18b] the terms bounded by Γ).

We note that the length of the time interval, c, is a universal constant for us. Indeed we have that ϕ is the identity (since in the notation of [CD18b], $b = \alpha$ for us), $\lambda = \frac{1}{4c_0}$ (this is the strong convexity constant for the Lagrangian which was assumed for us in Assumption 5), and L is the sum of the Lipschitz constants of $\partial_x H$ and $\partial_x V$ in space and in measure with respect to W_1, which is bounded by a universal constant due to Corollaries 3.3.6 and 3.4.4. ☐

Lemma 3.5.2. *If V is a classical solution to the master equation with regular data, then V is in fact as regular as the data.*

Proof. This is essentially the same as [GMMZ22, Proposition 6.3iii] and [MZ, Section 9.2]. ☐

Theorem 3.5.3. *Suppose that G, H are displacement monotone and satisfy Assumptions 4, 5, and 6.*

Then there is a unique global in time classical solution to the master equation (3.1.2).

Proof. We first prove uniqueness. Let $V, \tilde{V}$ be two classical solutions to the master equation. Because of the short time well-posedness of the system (3.4.1) we obtain that $\partial_x V = \partial_x \tilde{V}$. We will now use (3.1.2) to show that $V = \tilde{V}$. Let $u, \tilde{u}$ be the value functions given by (3.1.3) and $\rho, \tilde{\rho}$ be the associated solutions of the second equation of (3.1.2). Since $\partial_x V = \partial_x \tilde{V}$ we have that $\partial_x u = \partial_x \tilde{u}$ and so from the second equation of (3.1.2) we obtain that $\rho = \tilde{\rho}$. It now follows from Lemma 3.3.1 that $V = \tilde{V}$.

Next we prove existence. We will repeatedly apply the short time existence result Lemma 3.5.1. We let V_0 be a short time solution to the master equation on $[T - c, T]$. We can recursively define solutions V_k to the master equation by letting V_k be the short time solution on $[T - (k+1)\frac{c}{2}, T - k\frac{c}{2}]$ with the terminal condition $V_k(T - k\frac{c}{2}, \cdot) = V_{k-1}(T - k\frac{c}{2}, \cdot)$. Since c is a universal constant we will need only finitely many steps to cover the whole interval $[0, T]$. Because of the uniqueness proved above we have that the V_k's agree where their domains overlap and so we can stitch these together to obtain a classical solution to the master equation. ☐

CHAPTER 4

Canonical Transformations for Hamilton–Jacobi Equations

4.1 Introduction

4.1.1 Classical Mechanics and the Principle of Least Action

The Hamilton–Jacobi equation (HJE) in classical mechanics is deeply tied to the principle of least action, which provides a fundamental framework for describing the motion of a mechanical system. This principle, central to both Lagrangian and Hamiltonian mechanics, states that the actual trajectory of a system between two points in time is the one that extremizes the action functional.

Given a Lagrangian $L(x, \dot{x})$, the action $A[\gamma]$ along a trajectory $\gamma : [t_0, t_1] \to \mathbb{R}^d$ is defined as:

$$A[\gamma] = \int_{t_0}^{t_1} L(\gamma(s), \dot{\gamma}(s)) \, \mathrm{d}s.$$

The principle of least action asserts that the trajectory $\gamma(s)$ followed by the system is the one that makes the action stationary, leading to the Euler–Lagrange equations, which describe the system's dynamics.

However, an alternative and equally powerful approach to this problem is through the Hamiltonian formalism, where the system's dynamics are encoded in the Hamilton–Jacobi equation. To see how this arises, consider the value function $u(t, x)$, which represents the minimum action from time t to a fixed final time T, for a particle starting at position x at time t. This function can be written as:

$$u(t,x) = \inf_{\gamma:\gamma(t)=x} \int_t^T L(\gamma(s),\dot{\gamma}(s))\,\mathrm{d}s + G(\gamma(T)),$$

where $G(x)$ is a terminal cost or boundary condition at $t = T$. The function $u(t,x)$ gives the minimal accumulated action, or cost, of traveling from the point x at time t to the endpoint at time T.

4.1.1.1 Dynamic Programming and the Hamilton–Jacobi Equation

The value function $u(t,x)$ satisfies a fundamental property known as the *dynamic programming principle* (DPP). According to the DPP, the minimum action over the time interval $[t,T]$ can be decomposed into the minimum action over the shorter interval $[t,t+\delta t]$ plus the minimum action from $t+\delta t$ to T. Formally, this gives:

$$u(t,x) = \inf_{v\in\mathbb{R}^d} \left\{ \int_t^{t+\delta t} L(x,v)\,\mathrm{d}s + u(t+\delta t, x+v\delta t) \right\}.$$

Expanding $u(t+\delta t, x+v\delta t)$ using a first-order Taylor expansion yields:

$$u(t+\delta t, x+v\delta t) \approx u(t,x) + \delta t \left(\frac{\partial u}{\partial t} + v\cdot\nabla_x u \right).$$

Substituting this back into the dynamic programming principle and simplifying, we obtain the following expression for the minimal action over an infinitesimal time step:

$$u(t,x) = \inf_{v\in\mathbb{R}^d} \left\{ \delta t \left(L(x,v) + v\cdot\nabla_x u + \frac{\partial u}{\partial t} \right) + u(t,x) \right\}.$$

For this to hold, the term inside the brackets must vanish as $\delta t \to 0$, which leads directly to the Hamilton–Jacobi equation:

$$\frac{\partial u}{\partial t} + H\left(x, \nabla_x u\right) = 0,$$

where the Hamiltonian $H(x,p)$ is defined as the Legendre transform of the Lagrangian:

$$H(x,p) = \sup_{v\in\mathbb{R}^d} \left\{ p\cdot v - L(x,v) \right\}.$$

Thus, the Hamilton–Jacobi equation governs the evolution of the value function $u(t, x)$, encoding the minimal action required to reach the final state as time evolves. In this way, the HJE arises naturally from the principle of least action and the dynamic programming principle.

4.1.1.2 Connection to Euler–Lagrange Equations

The Hamilton–Jacobi equation provides an alternative formulation of classical mechanics that is fully equivalent to the Euler–Lagrange equations. While the Euler–Lagrange equations describe the explicit dynamics of the system in terms of the second-order differential equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}}\right) = \frac{\partial L}{\partial x},$$

the Hamilton–Jacobi equation focuses on the evolution of the value function $u(t, x)$, which encodes the system's entire future dynamics. By solving the HJE, one indirectly solves for the trajectory $\gamma(s)$ that minimizes the action, thereby determining the motion of the system.

In cases where the Hamiltonian $H(x, p)$ is integrable, the HJE can be solved explicitly, and the value function $u(t, x)$ provides a direct path to understanding the system's behavior. In more complex cases, where the HJE may not admit an explicit solution, the equation still offers valuable insights into the qualitative behavior of the system, including the formation of singularities and the presence of conserved quantities.

4.1.1.3 Relation to Hamilton's Equations

The Hamilton–Jacobi equation also provides an alternative way to derive Hamilton's equations of motion:

$$\dot{x} = \frac{\partial H}{\partial p}, \quad \dot{p} = -\frac{\partial H}{\partial x}.$$

In fact, if $u(t,x)$ is a solution to the Hamilton–Jacobi equation, the quantities

$$p_i = \frac{\partial u}{\partial x_i}$$

automatically satisfy Hamilton's equations. Thus, finding the solution to the HJE is equivalent to solving the equations of motion for the system.

This formalism is particularly advantageous for systems with multiple degrees of freedom, where the direct solution of Hamilton's equations may be complex. By solving the HJE, one reduces the problem to finding a suitable generating function, which encodes the entire dynamics of the system. In integrable systems, where there exist as many conserved quantities as degrees of freedom, the HJE can be solved by separation of variables, leading to a complete solution of the system.

4.1.2 Context on Coordinate Transformations in HJEs

The idea of using coordinate transformations to study Hamiltonian systems is well-known in analytic mechanics. For example, see [Arn89, Chapter 9] for an introduction to canonical transformations, and [Arn89, Appendix B] for a discussion on the role of canonical transformations in the study of Hamiltonian systems from the perspective of symplectic geometry.

4.1.3 Setup and Notation

For given data $H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and $G : \mathbb{R}^d \to \mathbb{R}$ and time horizon $T > 0$ let us consider the following Cauchy problem associated to the Hamilton–Jacobi–Bellman equation

$$\begin{cases} \partial_t u(t,x) + H(x, \partial_x u) = 0, & (t,x) \in (0,T) \times \mathbb{R}^d, \\ u(T,x) = G(x), & x \in \mathbb{R}^d, \end{cases} \tag{4.1.1}$$

For convenience we will assume that $H \in C^2(\mathbb{R}^d \times \mathbb{R}^d)$ and $G \in C^2(\mathbb{R}^d)$, and the second derivatives of these functions are uniformly bounded. Suppose that H is convex in its second variable, so that this can be seen as the Legendre–Fenchel transform of a Lagrangian function, i.e. we have

$$H(x,p) = \sup_{v\in\mathbb{R}^d}\{p\cdot v - L(x,-v)\}, \tag{4.1.2}$$

for some $L \in C^2(\mathbb{R}^d \times \mathbb{R}^d)$ given. For convenience, we also assume that L is convex in its second variable. In this case (4.1.1) corresponds to a variational problem. Indeed, it is well-known that under suitable assumptions we have that the value function

$$u(t,x) := \inf_{\gamma:[t,T]\to\mathbb{R}^d:\gamma(t)=x} \int_t^T L(\gamma(x),\dot\gamma(s))\mathrm{d}s + G(\gamma(T)) \tag{4.1.3}$$

is the unique viscosity solution to (4.1.1). This solution is locally Lipschitz continuous and locally semi-concave with a linear modulus of continuity (cf. [CS04, Theorem 7.4.12, Theorem 7.4.14]).

It is well-known, however, that the unique viscosity solution u in general develops singularities in finite time, even if H and G are smooth. Fine properties of sets of singularities of viscosity solutions have been studied extensively in the literature. Probably the first results on this topic were obtained in [CS87]. For a non-exhaustive list of further works dealing with singularities of solutions we refer the reader to [CS89, AC99, AC99, Yu06, CMS97, CY09, CF91, CF14] and to the review paper [CC21]. Singularity formation is equivalent to the non-uniqueness of optimal trajectories in the variational problem (4.1.3) (cf. [CS04]).

Therefore, if one is able to ensure uniqueness of optimizers in (4.1.3), this results in the differentiability of the value function, hence in the existence of a unique classical solution to (4.1.1). This is precisely the case if L is jointly convex and G is convex, when the dynamics in the control problem is linear. This implies that $u(t,\cdot)$ inherits the convexity and thus it becomes a $C^{1,1}_{loc}$ classical solution for arbitrary time horizon. This fact and related properties are classical and well documented in the literature, see for instance [CS04, Corollary 7.2.12] and [BE84, GR02, Goe05a, Goe05b, Roc70c, Roc70a].

As evidenced by the aforementioned works, the global existence of classical solutions is more of an exception than the rule in the theory of HJB equations. To the best of our knowledge, beyond the fully convex regime described above, there are no alternative sufficient conditions on the data (H, G) (or (L, G)) which would result in global in time classical well-posedness theory for (4.1.1) in the class $C^{1,1}_{loc}$.

In this chapter we show that a special class of linear canonical transformations can reveal new global well-posedness theories. Let us describe the philosophy behind our approach. Let $\alpha \in \mathbb{R}$ be given. Then the transformation

$$\mathbb{R}^d \times \mathbb{R}^d \ni (x, p) \mapsto (x, p - \alpha x)$$

is a so-called canonical transformation on the phase space, which preserves the structure of Hamilton's equations. Such transformations are well-known in classical mechanics (cf. [Arn89]). We will make use of the following definitions.

$$H_\alpha(x, p) := H(x, p - \alpha x) \tag{4.1.4}$$

and

$$G_\alpha(x) := G(x) + \frac{\alpha}{2}|x|^2. \tag{4.1.5}$$

Because of the nature of this transformation, we can state the first result of the chapter.

Theorem 4.1.1. *Let $\alpha \in \mathbb{R}$. Then u is a classical solution to* (4.1.1) *with data (H, G) in $(0, T) \times \mathbb{R}^d$, if and only if $u_\alpha : (0, T) \times \mathbb{R}^d$, defined as*

$$u_\alpha(t, x) := u(t, x) + \frac{\alpha}{2}|x|^2,$$

is a classical solution to (4.1.1) *on $(0, T) \times \mathbb{R}^d$ with data (H_α, G_α).*

This theorem has two immediate consequences. First, if we have a global well-posedness theory for (4.1.1) in the class $C^{1,1}_{loc}$ with data (H, G), we obtain a whole one parameter family of global well-posedness theories in $C^{1,1}_{loc}$ with data $(H_\alpha, G_\alpha)_{\alpha \in \mathbb{R}}$. Second, if we are able to find one real number $\alpha \in \mathbb{R}$ such that (4.1.1) is globally well-posed for the data (H_α, G_α), then the original problem with data (H, G) must also be globally well-posed.

It turns out that this second consequence will be the one revealing genuine new global in time well-posedness theories for (4.1.1) in the class $C^{1,1}_{loc}$. Therefore, as our second main result (formulated in Theorem 4.3.4 below), we have identified sufficient conditions on (H, G) which imply that for some precise $\alpha \in \mathbb{R}$, the transformed data (H_α, G_α) (or the corresponding (L_α, G_α)) fall into the well-known fully convex regime, and therefore this gives the global well-posedness of (4.1.1) in $C^{1,1}_{loc}$ with the original data (H, G). A direct corollary of our main results can be summarized as follows.

Corollary 4.1.2. *Let $H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ and $G : \mathbb{R}^d \to \mathbb{R}$ be C^2 functions with uniformly bounded second order derivatives. Suppose furthermore that $H(x, \cdot)$ is strongly convex, uniformly in x.*

Then, we have the followings.

1. *There exist a constant $C > 0$ depending only on $\|D^2G\|_\infty$, $\|D^2H\|_\infty$, and the lower bound on $\partial_{pp}H$ such that (4.1.1) with data $(\tilde{H}, G)$, where*

$$\tilde{H}(x,p) := H(x,p) + \alpha x \cdot p,$$

is globally well-posed in the class $C^{1,1}_{loc}([0,T] \times \mathbb{R}^d)$, for any $T > 0$, whenever $\alpha > C$.

2. *There exist a constant $C > 0$ depending only on $\|D^2G\|_\infty$, $\|D^2H\|_\infty$, and the lower bound on $\partial_{pp}H$ such that (4.1.1) with data $(\tilde{H}, G)$, where*

$$\tilde{H}(x,p) := H(x,p) - \alpha \frac{|x|^2}{2},$$

is globally well-posed in the class $C^{1,1}_{loc}([0,T] \times \mathbb{R}^d)$, for any $T > 0$, whenever $\alpha > C$.

Remark 4.1.3. We see that suitably modifying the existing data of a Hamilton–Jacobi–Bellman equation can 'convexify' the problem, and in turn this leads to a global in time classical well-posedness theory. Corollary 4.1.2 shows that this procedure can be done not only by adding the term $(x, p) \mapsto -\alpha\frac{|x|^2}{2}$ to the Hamiltonian, but also by adding $(x, p) \mapsto \alpha x \cdot p$ to H, for a suitably chosen α.

Remark 4.1.4. The attentive reader will notice that we are only using 'upper triangular' canonical transformations, i.e. transformations of the form $(x, p) \mapsto (x, p - \alpha x)$. The reason for this is that in order for a system of ODEs (in the (X_s, P_s) unknowns) to be the characteristic equations of a HJB equation not only do they need to have a Hamiltonian system structure but the boundary conditions must be of a particular form. Specifically to preserve the structure of boundary condition $X_0 = x_0$ we need to have the transformation to be upper triangular. From here one could imagine taking a transform $(x, p) \mapsto (x, p - Ax)$ for some constant matrix A. In order to preserve the structure of the condition $P_T = \nabla G(X_T)$ (specifically that the right-hand side is the gradient of a function) we need that A is symmetric. The choice of $A = \alpha I$ is taken for simplicity and the same arguments should work with any symmetric matrix A.

Remark 4.1.5. Although we illustrate this canonical transformation technique to obtain a classical global well-posedness theory for the HJB equation, the approach works exactly the same for any other features of HJB equations. For example, if one had a result on the structure of the shocks for a HJB equation with data (H, G) (e.g. for instance, the points of non-differentiability lie on a smooth curve) then the same result would hold for the transformed data (H_α, G_α) (in fact the non-differentiability points would be the exact same).

We finish this introduction with some concluding remarks.

- In this chapter, for the simplicity of the exposition, we choose to consider only a simple class of linear canonical transformations of the form $\mathbb{R}^d \times \mathbb{R}^d \ni (x, p) \mapsto (x, p - \alpha x)$. However, our approach would work for a class of more general transformations of the form

  $$\mathbb{R}^d \times \mathbb{R}^d \ni (x, p) \mapsto (x, p - \nabla\varphi(x)),$$

  for suitable potential functions $\varphi : \mathbb{R}^d \to \mathbb{R}$.

- Our approach works using only a Hamiltonian perspective, therefore, in particular by working purely with Hamiltonian systems, we believe that similar results could be proven in the case of Hamiltonians which are not necessarily convex in the momentum variable. Again, for simplicity of the exposition, we do not pursue this direction here.

- Canonical transformations are well understood in the case of more general Hamiltonian systems on symplectic manifolds, as these are symplectomorphisms on the cotangent bundle (cf. [Arn89]). Although we consider only the Euclidean setting here, we believe that our ideas could imply well-posedness theories for Hamilton–Jacobi–Bellman equations in more general geometric frameworks as well.

- It turns out the the canonical transformations that we have considered in this chapter reveal new deep well-posedness theories in a particular infinite dimensional setting, namely for the master equation in Mean Field Games. These results are detailed in the following chapter.

The rest of the chapter contains two short sections. In Section 4.2, for pedagogical reasons, we detail the role of the specific canonical transformations from the Lagrangian perspective. Section 4.3 contains our main results, and this is written purely from the Hamiltonian perspective.

4.2 Canonical transformations and classical solutions from the Lagrangian perspective

For $t \in (0,T)$, consider the functional $\mathcal{F}_t : C^1((t,T);\mathbb{R}^d) \to \mathbb{R}$ defined as

$$\mathcal{F}_t(\gamma) := \int_t^T L(\gamma(x), \dot\gamma(s))\mathrm{d}s + G(\gamma(T)).$$

Furthermore, we define the set of admissible curves as

$$\mathrm{Adm}_{t,x} := \{\gamma \in C^1((t,T);\mathbb{R}^d) : \gamma(t) = x\}.$$

Using this functional, one has

$$u(t,x) := \inf_{\gamma \in \mathrm{Adm}_{t,x}} \mathcal{F}_t(\gamma).$$

Differentiability of solutions to (4.1.1) is deeply linked to the uniqueness of minimizers in the optimal control problem (4.1.3). In particular, it is well-known that the convexity of the functional $\gamma \mapsto \mathcal{F}(\gamma)$ would imply that $u(t,\cdot)$ is convex, which in turn would further implies that u is a classical solution to (4.1.1) in the class $C^{1,1}_{loc}([0,T]\times\mathbb{R}^d)$ (see [CS04, Theorem 7.4.13]). The convexity of $\gamma \mapsto \mathcal{F}(\gamma)$ can be guaranteed by the joint convexity of L and the convexity of G.

However since the endpoint in the optimization problem, i.e. $\gamma(t) = x$, was fixed we could have just as well considered the functional

$$\mathcal{F}_{t,\alpha}(\gamma) := \mathcal{F}_t(\gamma) + \frac{\alpha}{2}|x|^2 = \mathcal{F}_t(\gamma) + \frac{\alpha}{2}|\gamma(t)|^2,$$

for any $\alpha \in \mathbb{R}$. In this case, one would simply have

$$u(t,x) + \frac{\alpha}{2}|x|^2 := \inf_{\gamma \in \mathrm{Adm}_{t,x}} \mathcal{F}_{t,\alpha}(\gamma). \tag{4.2.1}$$

Following a standard idea in classical mechanics we rewrite this new term as the integral of its time derivative and an initial term to get

$$\begin{aligned}\mathcal{F}_{t,\alpha}(\gamma) &= \int_t^T L(\gamma(x),\dot\gamma(s)) - \frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{\alpha}{2}|\gamma(s)|^2\right)\mathrm{d}s + \frac{\alpha}{2}|\gamma(T)|^2 + G(\gamma(T))\\ &= \int_t^T L(\gamma(x),\dot\gamma(s)) - \alpha\langle\gamma(s),\dot\gamma(s)\rangle \mathrm{d}s + G(\gamma(T)) + \frac{\alpha}{2}|\gamma(T)|^2\end{aligned}$$

Notice now that it is possible to not have convexity of

$$\gamma \mapsto \mathcal{F}_t(\gamma),$$

but have convexity of

$$\gamma \mapsto \mathcal{F}_{t,\alpha}(\gamma),$$

for some $\alpha \in \mathbb{R}$, even though

$$\inf_{\gamma \in \mathrm{Adm}_{t,x}} \mathcal{F}_t(\gamma) \quad and \quad \inf_{\gamma \in \mathrm{Adm}_{t,x}} \mathcal{F}_{t,\alpha}(\gamma)$$

are the same problems, in that the optimal values differ by the constant $\frac{\alpha}{2}|x|^2$ and in particular they have the same minimizers.

Therefore, it turns out that such transformations could reveal hidden convexity structures on the data, which were not straightforward in the original setting of the problem, and in particular (4.1.1) would be well-posed in the class $C^{1,1}_{loc}([0,T] \times \mathbb{R}^d)$, if $\gamma \mapsto \mathcal{F}_{t,\alpha}$ is convex even if we did not have convexity of $\gamma \mapsto \mathcal{F}_t$.

We also have the opposite situation, i.e. when $\gamma \mapsto \mathcal{F}_t$ is convex, yet $\gamma \mapsto \mathcal{F}_{t,\alpha}$ is not convex. As the convexity properties of the functionals can be characterized by the convexity of the Lagrangian and final data, it is natural to define the following quantities. For $\alpha \in \mathbb{R}$, let $L_\alpha : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be defined as

$$L_\alpha(x, v) := L(x, v) - \alpha x \cdot v$$

and $G_\alpha : \mathbb{R}^d \to \mathbb{R}$, defined as

$$G_\alpha(x) := G(x) + \frac{\alpha}{2}|x|^2.$$

Based on the previous discussion, we can formulate the following proposition.

Proposition 4.2.1. *We have the following.*

(i) Let $G : \mathbb{R}^d \to \mathbb{R}$ be convex and let $L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ be jointly convex, and suppose furthermore that both G and L have bounded second derivatives. Then, there exists $\alpha_0 \in \mathbb{R}$ such that L_α is not jointly convex and G_α is not convex for any $\alpha < \alpha_0$.

(ii) There exist $L : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ not jointly convex and $G : \mathbb{R}^d \to \mathbb{R}$ non-convex, such that there for a suitable $\alpha \in \mathbb{R}$, L_α becomes jointly convex and G_α becomes convex.

Proof. (i) Let $f : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ defined as $f(v, x) = v \cdot x$. We see that -1 is an eigenvalue of $D^2 f$ (the eigenvector is the $(1, \ldots, 1, -1, \ldots, -1)$). Hence for $\alpha < \hat{\alpha_0} := -\|D^2 L\|_\infty$ we have that L_α is not jointly convex. Now let $\alpha_0 := \min\{\hat{\alpha}_0, -\|\partial_{xx}G\|_{L^\infty}\}$, and then the result follows.

(ii) In the previous point we have constructed L_α and G_α that are non-convex, but L and G were convex. Now if we apply the same transformation on these new functions with constant $-\alpha$, i.e. $(L_\alpha)_{-\alpha}$ and $(G_\alpha)_{-\alpha}$ we get back to the original functions which were convex. The statement follows. □

The transformations on L, as describe above, translate naturally to the Hamiltonian H. Indeed, we can see that H_α corresponding to L_α is defined as in 4.1.4.

It is important to notice that the previous transformation preserves the Hamiltonian structure and the HJB equation. This is what leads precisely to Theorem 4.1.1, whose proof is straightforward and we present it below.

Proof of Theorem 4.1.1. This result readily follows from the representation formula (4.2.1). Alternatively, direct computation yields

$$\partial_t u_\alpha(t,x) = \partial_t u(t,x), \quad and, \partial_x u_\alpha(t,x) = \partial_x u(t,x) + \alpha x,$$

and so

$$-\partial_t u_\alpha(t,x) + H_\alpha(x, \partial_x u_\alpha(t,x)) = -\partial_t u(t,x) + H(x, \partial_x u(t,x) + \alpha x - \alpha x) = 0,$$

and

$$u_\alpha(T,x) = G_\alpha(x).$$

The result follows. □

Remark 4.2.2. Because of the representation formula (4.2.1), the previous result clearly holds true for viscosity solutions as well.

4.3 Canonical transformations and classical solutions from the perspective of Hamiltonian systems

Based on [Roc70b, Theorem 33.1], we can formulate the following result.

Lemma 4.3.1. *$H : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, defined in (4.1.2) is concave-convex (i.e. $H(\cdot, p)$ is convex for all $p \in \mathbb{R}^d$ and $H(x, \cdot)$ is convex for all $x \in \mathbb{R}^d$) if and only L is jointly convex.*

From this lemma we see that the global existence of classical solutions to the Hamilton–Jacobi equation (4.1.1) in the class $C^{1,1}_{loc}$, from the Hamiltonian point of view, is intimately linked to the concave-convex properties of H and convexity of the final condition G.

Definition 4.3.2. For a square matrix $A \in \mathbb{R}^{m\times m}$, we define the symmetric matrix

$$\operatorname{Re} A := \frac{1}{2}(A + A^\top).$$

For a symmetric matrix $A \in \mathbb{R}^{m\times m}$, we denote by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ its smallest and largest eigenvalues, respectively.

Lemma 4.3.3. *Suppose that*

$$\left(w^\top \operatorname{Re} \partial_{xp} H(x,p) w\right)^2 - \left(w^\top \partial_{pp} H(x,p) w\right)\left(w^\top \partial_{xx} H(x,p) w\right) \geq 0, \quad \forall w \in \mathbb{R}^d, \quad \forall x, p \in \mathbb{R}^d \times \mathbb{R}^d. \tag{4.3.1}$$

Define

$$\alpha := \inf_{\substack{(x,p,w)\in\mathbb{R}^{3d} \\ \|w\|=1}} \frac{w^\top \operatorname{Re} \partial_{xp} H(x,p) w + \sqrt{\left(w^\top \operatorname{Re} \partial_{xp} H(x,p) w\right)^2 - \left(w^\top \partial_{pp} H(x,p) w\right)\left(w^\top \partial_{xx} H(x,p) w\right)}}{w^\top \partial_{pp} H(x,p) w}. \tag{4.3.2}$$

Suppose that $x \mapsto G(x) + \alpha \frac{|x|^2}{2}$ is convex and

$$\alpha \geq \sup_{\substack{(x,p,w)\in\mathbb{R}^{3d}\\ \|w\|=1}} \frac{w^\top \operatorname{Re}\partial_{xp}H(x,p)w - \sqrt{\left(w^\top \operatorname{Re}\partial_{xp}H(x,p)w\right)^2 - \left(w^\top\partial_{pp}H(x,p)w\right)\left(w^\top\partial_{xx}H(x,p)w\right)}}{w^\top\partial_{pp}H(x,p)w}. \tag{4.3.3}$$

Then the Hamilton–Jacobi equation (4.1.1) *with data* (H, G) *is globally well-posed in the class* $C^{1,1}_{loc}([0,T]\times\mathbb{R}^d)$.

Proof. Using (4.1.4) and (4.1.5) we define H_α and G_α with the particular choice of α given in the statement. We see that G_α is convex. Also, we compute for any $w \in \mathbb{R}^d$ and any $(x,p) \in \mathbb{R}^d \times \mathbb{R}^d$

$$\begin{aligned} &w^\top\partial_{xx}H_\alpha(x,p)w \\ &= w^\top\partial_{xx}H(x,p-\alpha x)w - 2\alpha w^\top \operatorname{Re}(\partial_{xp}H(x,p-\alpha x))w + \alpha^2 w^\top\partial_{pp}H(x,p-\alpha x)w \end{aligned}$$

This expression is a quadratic polynomial in α with positive leading coefficient. The conditions of the theorem assure that this polynomial is non-positive at α, i.e.

$$\begin{aligned} &\frac{w^\top \operatorname{Re}\partial_{xp}H(x,p)w - \sqrt{(w^\top \operatorname{Re}\partial_{xp}H(x,p)w)^2 - (w^\top\partial_{pp}H(x,p)w)(w^\top\partial_{xx}H(x,p)w)}}{w^\top\partial_{pp}H(x,p)w} \\ &\leq \alpha \\ &\leq \frac{w^\top \operatorname{Re}\partial_{xp}H(x,p)w + \sqrt{(w^\top \operatorname{Re}\partial_{xp}H(x,p)w)^2 - (w^\top\partial_{pp}H(x,p)w)(w^\top\partial_{xx}H(x,p)w)}}{w^\top\partial_{pp}H(x,p)w}, \end{aligned}$$

for all $(x,p) \in \mathbb{R}^d \times \mathbb{R}^d$ and for all $w \in \mathbb{R}^d$.

In particular H_α is concave in x. Furthermore, this particular transformation does not change the convexity of H_α in the p-variable, as $\partial_{pp}H_\alpha(x,p) = \partial_{pp}H(x,p-\alpha x)$. The thesis of the lemma follows by Lemma 4.3.1 and [CS04, Theorem 7.4.13]. □

As a consequence of this lemma, we can formulate the following result.

Theorem 4.3.4. *We define the following quantities*

$$\lambda_0 := \inf_{(x,p)\in\mathbb{R}^d\times\mathbb{R}^d} \lambda_{\min}\left(\operatorname{Re}\partial_{xp}H(x,p)\right),$$

$$\lambda_H := \sup_{(x,p)\in\mathbb{R}^d\times\mathbb{R}^d} \lambda_{\max}\left(\partial_{xx}H(x,p)\right)$$

and

$$\lambda_G := \inf_{x\in\mathbb{R}^d} \lambda_{\min}\left(\partial_{xx}G(x)\right).$$

Suppose that

$$\lambda_0^2 \geq \|\partial_{pp}H\|_\infty \lambda_H \quad \text{and that} \quad \lambda_0 + \sqrt{\lambda_0^2 - \|\partial_{pp}H\|_\infty\lambda_H} + \|\partial_{pp}H\|_\infty\lambda_G \geq 0.$$

Furthermore assume that either $\lambda_H \leq 0$ or $\lambda_0 \geq 0$. Then the Hamilton–Jacobi equation (4.1.1) *is globally well-posed, for any $T > 0$, in the class $C^{1,1}_{loc}([0,T]\times\mathbb{R}^d)$.*

Proof. We verify the assumptions of Lemma 4.3.3. First, let us consider the inequality (4.3.1).

If $\lambda_H \leq 0$ then (4.3.1) is fulfilled immediately.

If $\lambda_H > 0$, but $\lambda_0 \geq 0$ we have the following. By definition of λ_0, $w^\top \operatorname{Re}\partial_{xp}H(x,p)w \geq \lambda_0$, for all $(x,p)\in\mathbb{R}^d\times\mathbb{R}^d$ and for all $w\in\mathbb{R}^d$. Since the right side is non-negative we can square to obtain

$$(w^\top \operatorname{Re}\partial_{xp}H(x,p)w)^2 \geq \lambda_0^2 \geq \|\partial_{pp}H\|_\infty\lambda_H,$$

which implies that

$$(w^\top \operatorname{Re}\partial_{xp}H(x,p)w)^2 \geq (w^\top\partial_{pp}H(x,p)w)(w^\top\partial_{xx}H(x,p)w).$$

Thus, (4.3.1) follows.

We verify the other assumptions in the statement of Lemma 4.3.3. Let α be defined as in (4.3.2). Just as before, we distinguish two cases.

Case 1. $\lambda_H \leq 0$.

In this case we have that $(w^\top \partial_{pp} H(x,p) w)(w^\top \partial_{xx} H(x,p) w) \leq 0$, for all $(x,p) \in \mathbb{R}^d \times \mathbb{R}^d$ and for all $w \in \mathbb{R}^d$. Therefore

$$\begin{aligned}
\alpha &\geq 0 \\
&\geq \frac{w^\top \operatorname{Re} \partial_{xp} H(x,p) w - \sqrt{(w^\top \operatorname{Re} \partial_{xp} H(x,p) w)^2 - (w^\top \partial_{pp} H(x,p) w)(w^\top \partial_{xx} H(x,p) w)}}{w^\top \partial_{pp} H(x,p) w},
\end{aligned}$$

for all $(x,p) \in \mathbb{R}^d \times \mathbb{R}^d$ and for all $w \in \mathbb{R}^d$. This implies (4.3.3).

Furthermore, we have

$$\begin{aligned}
\alpha &= \\
&\inf_{\substack{(x,p,w) \in \mathbb{R}^{3d} \\ \|w\|=1}} \frac{w^\top \operatorname{Re} \partial_{xp} H(x,p) w + \sqrt{\left(w^\top \operatorname{Re} \partial_{xp} H(x,p) w\right)^2 - \left(w^\top \partial_{pp} H(x,p) w\right)\left(w^\top \partial_{xx} H(x,p) w\right)}}{w^\top \partial_{pp} H(x,p) w} \\
&\geq \inf_{\substack{(x,p,w) \in \mathbb{R}^{3d} \\ \|w\|=1}} \frac{w^\top \operatorname{Re} \partial_{xp} H(x,p) w + \sqrt{\left(w^\top \operatorname{Re} \partial_{xp} H(x,p) w\right)^2 - \|\partial_{pp} H\|_\infty \left(w^\top \partial_{xx} H(x,p) w\right)}}{\|\partial_{pp} H\|_\infty}.
\end{aligned}$$

where in the last inequality we have used that the function $f : \{(a,b,c) : \ c \geq 0, \ b \leq 0, \ a^2 \geq bc\} \to \mathbb{R}$ defined as $f(a,b,c) = \frac{a + \sqrt{a^2 - bc}}{c}$ is decreasing in c.

Continuing we have

$$\begin{aligned}
\alpha &\geq \\
&\inf_{\substack{(x,p,w) \in \mathbb{R}^{3d} \\ \|w\|=1}} \frac{w^\top \operatorname{Re} \partial_{xp} H(x,p) w + \sqrt{\left(w^\top \operatorname{Re} \partial_{xp} H(x,p) w\right)^2 - \|\partial_{pp} H\|_\infty \left(w^\top \partial_{xx} H(x,p) w\right)}}{\|\partial_{pp} H\|_\infty} \\
&\geq \inf_{\substack{(x,p,w) \in \mathbb{R}^{3d} \\ \|w\| = 1}} \frac{w^\top \operatorname{Re} \partial_{xp} H(x,p) w + \sqrt{(w^\top \operatorname{Re} \partial_{xp} H(x,y) w)^2 - \|\partial_{pp} H\|_\infty \lambda_H}}{\|\partial_{pp} H\|_\infty} \\
&\geq \frac{\lambda_0 + \sqrt{\lambda_0^2 - \|\partial_{pp} H\|_\infty \lambda_H}}{\|\partial_{pp} H\|_\infty}
\end{aligned}$$

where the last inequality is because the function $f : \{(a,b) : \ b \leq 0\} \to \mathbb{R}$ defined as $f(a,b) = a + \sqrt{a^2 - b}$ is increasing in a. From this, by the assumptions of this theorem it follows that $x \mapsto G(x) + \alpha \frac{|x|^2}{2}$ is convex.

Case 2. $\lambda_0 \geq 0$. We notice that without loss of generality, we may assume also that the inequality $\lambda_H \geq 0$ takes place.

We have

$$
\begin{aligned}
\alpha = & \\
& \inf_{\substack{(x,p,w)\in\mathbb{R}^{3d}\\ \|w\|=1}} \frac{w^\top \operatorname{Re}\partial_{xp}H(x,p)w + \sqrt{\left(w^\top \operatorname{Re}\partial_{xp}H(x,p)w\right)^2 - \left(w^\top\partial_{pp}H(x,p)w\right)\left(w^\top\partial_{xx}H(x,p)w\right)}}{w^\top\partial_{pp}H(x,p)w} \\
& \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (4.3.4) \\
\geq & \inf_{\substack{(x,p,w)\in\mathbb{R}^{3d}\\ \|w\|=1}} \frac{\lambda_0 + \sqrt{\lambda_0^2 - (w^\top\partial_{pp}H(x,p)w)\lambda_H}}{w^\top\partial_{pp}H(x,p)w} \\
\geq & \inf_{\substack{(x,p,w)\in\mathbb{R}^{3d}\\ \|w\|=1}} \frac{\lambda_0 + \sqrt{\lambda_0^2 - \|\partial_{pp}H\|_\infty\lambda_H}}{w^\top\partial_{pp}H(x,p)w} \\
\geq & \frac{\lambda_0 + \sqrt{\lambda_0^2 - \|\partial_{pp}H\|_\infty\lambda_H}}{\|\partial_{pp}H\|_\infty}
\end{aligned}
$$

where the last two inequalities follow from $\lambda_H \geq 0$ and $\lambda_0 \geq 0$ respectively. We notice also that in the previous chain of inequalities all the quantities under the square root are non-negative.

Furthermore, we see that $\lambda_0+\sqrt{\lambda_0^2 - \|\partial_{pp}H\|_\infty\lambda_H}+\|\partial_{pp}H\|_\infty\lambda_G \geq 0$ implies that $\alpha+\lambda_G \geq 0$ and so $x \mapsto G(x) + \alpha\frac{|x|^2}{2}$ is convex.

Next note that the function $f : \{(a,b) : \ a,b \geq 0,\ a^2 \geq b\} \to \mathbb{R}$ defined as $f(a,b) =$

$a - \sqrt{a^2 - b}$ is decreasing in a. Hence

$$\begin{aligned} &\frac{w^\top \operatorname{Re} \partial_{xp} H(x,p) w - \sqrt{(w^\top \operatorname{Re} \partial_{xp} H(x,p)w)^2 - (w^\top \partial_{pp} H(x,p) w)(w^\top \partial_{xx} H(x,p) w)}}{w^\top \partial_{pp} H(x,p) w} \\ &\leq \frac{\lambda_0 - \sqrt{\lambda_0^2 - (w^\top \partial_{pp} H(x,p) w)(w^\top \partial_{xx} H(x,p) w)}}{w^\top \partial_{pp} H(x,p) w} \\ &\leq \frac{\lambda_0 - \sqrt{\lambda_0^2 - (w^\top \partial_{pp} H(x,p) w)\lambda_H}}{w^\top \partial_{pp} H(x,p) w} \\ &\leq \frac{\lambda_0 - \sqrt{\lambda_0^2 - \|\partial_{pp} H\|_\infty \lambda_H}}{\|\partial_{pp} H\|_\infty} \end{aligned} \tag{4.3.5}$$

where the last inequality follows from the fact that the function $f : \{(a,b,c) : \ a,b,c \geq 0,\ a^2 \geq bc\} \to \mathbb{R}$, defined as $f(a,b,c) = \frac{a - \sqrt{a^2 - bc}}{c}$ is increasing in c. Combining (4.3.4) and (4.3.5) we can conclude that

$$\alpha \geq \sup_{\substack{(x,p,w) \in \mathbb{R}^{3d} \\ \|w\|=1}} \frac{w^\top \operatorname{Re} \partial_{xp} H(x,p) w - \sqrt{\big(w^\top \operatorname{Re} \partial_{xp} H(x,p) w\big)^2 - \big(w^\top \partial_{pp} H(x,p) w\big)\big(w^\top \partial_{xx} H(x,p) w\big)}}{w^\top \partial_{pp} H(x,p) w},$$

which completes the proof in this case.

□

From this theorem the proof of Corollary 4.1.2 is immediate.

Proof of Corollary 4.1.2. We see that if α is large enough then the assumptions of Theorem 4.3.4 are satisfied. □

CHAPTER 5

Hidden Monotinicity in Master Equation

5.1 Introduction

Mean field games (MFGs), first introduced in the seminal works of Lasry and Lions as well as Huang, Malhamé, and Caines (see [LL07, HMC06]), were motivated by the need to model strategic decision-making in large-scale systems with numerous rational agents, stemming from (stochastic) differential games. Since then, this theory has seen widespread success, both theoretically and in practical applications. For a thorough and relatively current overview of the field's progress from probabilistic and analytic perspectives, see [CD18a, CD18b, CP20b].

In his lecture series at Collège de France ([Lio12a]), Lions introduced what is now known as the master equation for MFGs. This equation is a nonlocal, nonlinear PDE of hyperbolic type set on $\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$, where $\mathbb{R}^d$ models an individual agent's state space, and $\mathscr{P}_2(\mathbb{R}^d)$ represents the set of Borel probability measures on $\mathbb{R}^d$ with finite second moments, encoding the distribution of agents. A major motivation for solving the master equation lies in its deep link between games with a finite, but large, number of agents and the associated MFG. Specifically, classical solutions to the master equation are instrumental in deriving quantitative rates of convergence for closed-loop Nash equilibria in finite-agent games as the number of agents approaches infinity.

The search for well-posedness theories for (1.1) has catalyzed an extensive program in the field. The non-local and infinite-dimensional nature of this PDE introduces considerable challenges. In particular, the lack of a comparison principle means that approaches relying on viscosity solutions, for instance, are not feasible. Consequently, there is potential for debate in defining suitable weak solutions, especially where uniqueness is not guaranteed. Nevertheless, there is clarity when it comes to classical solutions, and in this chapter, we will focus on classical solutions. Unless otherwise stated, the term well-posedness will refer to classical solutions. As with finite-dimensional conservation laws, global classical solutions generally require specific monotonicity conditions on data elements H and G, which also strongly relate to the uniqueness of MFG Nash equilibria.

Literature review on the well-posedness of master equations. To date, there have been different notions of monotonicity conditions proposed on the data H and G, which could serve as sufficient conditions for the global well-posedness theory of (1.1). The diversity and richness of these conditions are deeply related to the geometry under the lens of which we look at $\mathscr{P}_2(\mathbb{R}^d)$. For instance, $\mathscr{P}_2(\mathbb{R}^d)$ can be seen as a flat convex space, but it is natural to look at it also as a non-negatively curved infinite dimensional manifold, when equipped with suitable metrics. Historically, the so-called Lasry–Lions (LL) monotonicity condition was the first one, introduced already in the seminal work [LL07]. Geometrically, this is linked to the flat geometry, imposed on $\mathscr{P}_2(\mathbb{R}^d)$. When it comes to nonlocal Hamiltonians, this notion has been defined and exploited so far only for so-called separable Hamiltonians, i.e. the ones which have the structure

$$H(x,\mu,p) := H_0(x,p) - F(x,\mu), \quad \forall (x,\mu,p) \in \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d, \tag{5.1.1}$$

for some H_0 and F. An alternative monotonicity condition is the so-called displacement monotonicity condition, which does not require the separable structural assumption on H. This stems from the notion of displacement convexity, used widely in the context of optimal transport theory. Thus, this is linked to the curved geometry on $\mathscr{P}_2(\mathbb{R}^d)$. We now give a brief overview of the well-posedness theories for (1.1) in these settings and we also mention some alternative, more recently proposed notions of monotonicity conditions.

In [CD18b, Theorem 5.46] the authors have shown that the master equation (1.1) is globally well-posed if the data are LL monotone and possess additional regularity assumptions. Several other works provide similar conclusions. We refer to [CDLL19, Theorem 2.4.5] for the case when the physical space is the flat torus instead of $\mathbb{R}^d$ and to [CCD22, Theorems 56 and 58] to the case without common noise (i.e. $\beta_0 = 0$). We refer also to [JR23] for new results and clarifications regarding the results from [CDLL19]. However, [CD18b, Theorem 5.46] is the closest result for our purposes.

It is also important to mention that all these global well-posedness results in the context of Lasry–Lions monotonicity impose both the separable structure on the Hamiltonian and the presence of a non-degenerate idiosyncratic noise.

In the context of displacement monotonicity global in time well-posedness have been obtained chronologically as follows. [GM22a] provided this in the context of deterministic and potential (in particular $\beta = \beta_0 = 0$ and H separable) games (for similar results, see also [BGY24]). [GMMZ22] provided the first global in time well-posedness result in the case of non-separable displacement monotone Hamiltonians and non-degenerate idiosyncratic noise (i.e. $\beta \neq 0$). Finally, [BMM23a] provided the result in the case of degenerate idiosyncratic noise (i.e. $\beta = 0$) and compared to [GMMZ22], under lower level regularity assumptions on the data, and the weaker version of the displacement monotonicity condition on H.

Recently, in [MZ22b] and [MZ24] the authors have proposed a notion of anti-monotonicity condition on final data of master equations, which together with other sufficient structural conditions on the Hamiltonian resulted in the the global in time well-posedness of the master equation. We would like to emphasize that for this to hold, the anti-monotonicity condition on the final data has to be carefully chosen in line with the structural conditions on the Hamiltonian. As we show below, this framework can entirely be embedded into our main results under the umbrella of our newly proposed canonical transformation.

Several other recent developments have seen the light in the context of the well-posedness of MFG master equations. For a non-exhaustive list we refer to [AM23, Ber21, CCP23, CD24, GM22c, GM23].

Description of results. In this chapter our main objective is to present new global well-posedness theories for the master equation (1.1). The heart of our analysis consist of so-called canonical transformations which in particular reveal new perspectives on existing and new monotonicity conditions on the Hamiltonians and final data associated to (1.1), and in turn lead to new well-posedness theories. The values of the noise intensities, β, β_0 will not not be significant in our consideration, and our main results hold true also for degenerate problems, i.e. when $\beta = 0$ or $\beta_0 = 0$.

In classical Hamiltonian mechanics, canonical transformations are coordinate transformations on the phase space, which preserve the structure of Hamilton's equations. In symplectic geometry, canonical transforms are known as symplectomorphisms (where the phase space is a cotangent bundle and the symplectic form is the canonical 2-form). Since in our setting we are only concerned with Euclidean space we do not use the symplectic terminology. However, one could use symplectomorphisms to generate new well-posedness theories for Hamilton–Jacobi equations and the master equation in more general settings (i.e. when the underlying space is not Euclidean). We refer the reader to [Arn89] for a introduction to applications of symplectic geometry in classical mechanics. We refer also to our companion short note [BM24a], where we explain the regularization effect of such transformations in the case of deterministic finite dimensional HJB equations.

As the master equation has in particular a natural character arising from infinite dimensional Hamiltonian dynamics, we will show below, that such transformations play a deep role in obtaining new well-posedness theories for it.

Let us describe the driving idea behind our results. For Hamiltonians $H : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ and final data $G : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ we consider a family of prototypical linear canonical transformations as follows. Let $\alpha \in \mathbb{R}$ and define $H_\alpha : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$

and $G_\alpha : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ as

$$H_\alpha(x,\mu,p) := H(x,\mu,p-\alpha x) \;\; and \;\; G_\alpha(x,\mu) := G(x,\mu) + \frac{\alpha}{2}|x|^2. \tag{5.1.2}$$

In particular, this means that the corresponding canonical transformation has the form of

$$\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \ni (x,\mu,p) \mapsto (x,\mu,x-\alpha p).$$

This is a 'finite dimensional' transformation, as there is no change in the measure variable μ. Having defined these transformations, we can formulate our first main result.

Theorem 5.1.1. *Fix any $\alpha \in \mathbb{R}$. The master equation with data (H,G) is well-posed if and only if it is well-posed with data (H_α, G_α).*

The message of this theorem is that if one produces a well-posedness theory for the master equation, this will lead to a whole one parameter family of well-posedness theories, with the transformed data. A deeper consequence of this theorem is the opposite implication. Suppose that one is given the data (H,G). If one is able to find a suitable range of the parameter α such that (H_α, G_α) satisfies some well-known monotonicity conditions, then the problem with the original data must be well-posed. This second one will be the direction that we investigate in this chapter.

Fix $\alpha \in \mathbb{R}$. It is easy to see that G is LL monotone, if and only if G_α is LL monotone and the situation is the same for separable H. However, as we will show below, this phenomenon is much different in the displacement monotone regime. Therefore Theorem 5.1.1 has powerful applications in the context of displacement monotonicity but not for LL monotonicity.

In Theorem 5.2.8, we propose easily verifiable sufficient conditions on H to ensure that H_α is displacement monotone.

This theorem has an immediate consequence which can informally be formulated as follows. This result provides a new global well-posedness theory for the master equation.

Corollary 5.1.2. *Suppose that* $G : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ *and* $H : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ *are twice continuously differentiable with uniformly bounded second order derivatives. Suppose moreover that* H *is strongly convex in the* p*-variable.*

We have that there exists $C > 0$ *depending on second derivatives of* H *and* G *(but independent of* T*) so that if* $\alpha \geq C$ *then the master equation is globally well-posed with data* $(\tilde{H}, G)$*, where* $\tilde{H} : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ *is given by*

$$\tilde{H}(x, \mu, p) := H(x, \mu, p) + \alpha p \cdot x.$$

Hence even if we did not know that the original master equation was solvable, the modified master equation is solvable for α large enough. One can compare the Hamiltonian $\tilde{H}$ with the one in [MZ24, Example 7.2].

Remark 5.1.3. Corollary 5.1.2 has a deep message: if the Hamiltonian is such that $\partial_{xp} H$ is sufficiently large compared to other second order derivatives of H and the second order derivatives of G, then we have a global well-posedness theory for the master equation. Therefore $\partial_{xp} H$, and in particular adding suitable multiples of the function $(x, p, \mu) \mapsto p \cdot x$ to H have a 'regularization effect'. By carefully examining Lemma 5.2.6, we see that what is going on is that the $p \cdot x$ term is transformed into a multiple of $\frac{|x|^2}{2}$, which provides displacement monotonicity for the problem and hence regularizes the master equation. It is easy to see that adding a suitable multiple of the term $\frac{|x|^2}{2}$ to H produces displacement monotonicity. Clearly, these regularization effects are independent of the noise intensities.

Further implications of main results. Having our main results in hand, we have revisited some previous well-posedness results from the literature.

When G is displacement semi-monotone, then the well-posedness of (1.1) can be guaranteed if H_α is displacement monotone for sufficiently large α. It turns out that our characterization for this given in Proposition 5.2.6 coincides with the respective assumptions on H discovered recently in [MZ22b].

In the recent paper [MZ24], the authors proposed a notion of anti-monotonicity for final data G. They have described some heavy sufficient conditions on H and G which result in a global well-posedness theory of (1.1), if $\beta \neq 0$, and G is suitably anti-monotone. There was an emphasis on the fact that G needed to be 'sufficiently' anti-monotone.

It turns out that these well-posedness results from [MZ24], under the additional assumptions that H is strictly convex in the p-variable fall directly into the framework of the canonical transformations and they are an easy consequence of our main results, from Corollary 5.1.2. More precisely, first in Proposition 5.2.10 we show that if G is λ-anti-monotone, this implies that it is displacement semi-monotone with a constant which depends only on λ (in particular, the displacement semi-monotonicity constant is independent of the second derivative bounds of G). Having strong convexity of H in the p-variable, which has also bounded second derivatives allows us to use our Corollary 5.1.2. The Hamiltonian considered in [MZ24] has the form of

$$H(x, \mu, p) := H_0(x, \mu, p) + \langle A_0 p, x \rangle,$$

for some constant matrix $A_0 \in \mathbb{R}^{d \times d}$. This is slightly different than $\tilde{H}$ from our Corollary 5.1.2, but the term $\langle A_0 p, x \rangle$ has exactly the same role as $\alpha p \cdot x$ in our consideration. Therefore, for completeness, as our last contributions, in Proposition 5.2.14 and Remark 5.2.15 we show that the assumptions from the main theorem in [MZ24] essentially imply our assumptions. Furthermore, in the case of Hamiltonians which are strongly convex in the p-variable, our results need less and weaker assumption, and they hold true without the presence of a non-degenerate idiosyncratic noise. In particular, we demonstrate that the emphasis on the sufficient anti-monotonicity of G in [MZ24] is misleading, and this is not needed. Specifically, in [MZ24] it is remarked: "... we will need to require our data to be sufficiently anti-monotone in appropriate sense". However we will see that anti-monotonicty is not needed (as anti-monotonicity implies semi-monotonicity) and that [MZ24] has other other, more essential assumptions on H which are what really give the well-posedness result.

Some concluding remarks.

- For simplicity and transparency of our main ideas, in this chapter we have decided to focus only on linear canonical transformations of the form $\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \ni (x,\mu,p) \mapsto (x,\mu,x-\alpha p)$. Without much philosophical effort but with significant technical effort, one could consider canonical transformations of the form

$$\mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \ni (x,\mu,p) \mapsto (x,\mu,x-\nabla\varphi(x)),$$

where $\varphi : \mathbb{R}^d \to \mathbb{R}$ is any given smooth potential function, with bounded second derivatives. In the case of noise, this transformation would lead to the modified Hamiltonians and final data as

$$H_\varphi(x,\mu,p) := H(x,\mu,x-\nabla\varphi(x)) + \frac{\beta^2+\beta_0^2}{2}\Delta\varphi(x) \quad \textit{and} \quad G_\varphi(x,\mu) := G(x,\mu) + \varphi(x).$$

It is easy to see that Theorem 5.1.1 holds true if in its statement (H_α, G_α) is replaced with (H_φ, G_φ). However, in order to obtain new global well-posedness theory (in the case of potentially degenerate noise), we would need to have a 'convexifying regularization' on G_φ, which means that φ would need to be taken to be convex with sufficiently large Hessian eigenvalues. From this point of view, $\varphi(x) = \frac{\alpha}{2}|x|^2$ would be a natural choice, and this is why we have decided to reduce our study to this particular family of potentials.

We remark that in general Hamiltonians are only defined up to an additive constant. In classical mechanics, this is saying that we may pick any value to correspond to the 'zero energy'. In the presence of noise the attentive reader will notice that our H_α is not the same as the H_φ defined above, when $\varphi(x)$ is taken to be $\frac{\alpha}{2}|x|^2$. However, this is not an issue as the difference between the two is a constant. In particular, the two Hamiltonians are equivalent. Thus, we could have defined our H_α as $H_\alpha(x,\mu,p) := H(x,\mu,p-\alpha x) + \frac{(\beta^2+\beta_0^2)d}{2}\alpha$ which would then be the exact same as H_φ defined above, however this would introduce unnecessary notational clutter.

- In this chapter we have considered only ‘finite dimensional’ canonical transformations (where the measure component stayed fixed). These have proved to have a deep effect on new global well-posedness theories for the master equation. It is a very interesting, but seemingly challenging task to analyze truly infinite dimensional canonical transformations in the context of MFG master equations. In particular it seems that the infinite dimensional canonical transformations do not preserve the structure of MFG, they only preserve the structure of optimal control problems. In this we see a significant difference between games and variational problems.

Remark 5.1.4. If the Hamiltonian H has an associated Lagrangian with bounded second derivatives we must have that H is strongly convex in p. Similarly, the master equation only corresponds to a game, when H is convex in p. To the best of the authors knowledge there is no motivation for the master equation outside of this case.

We remark that if one is interested in the case of non-convex H in p then one can adapt our results by using the Hamiltonian system directly. We refer to the Lagrangian purely for pedagogical reasons and it is not needed for any technical reason. In particular our canonical transformation and main theorem, Theorem 5.1.1, holds regardless of the convexity of H in p.

5.2 Preliminaries and well-posedness theories for master equations

5.2.1 Some notations

In order to keep this discussion self-contained, let us recall some definitions and notations.

Let $p \geq 1$. Based on [AGS08], we recall that the p-Wasserstein between $\mu, \nu \in \mathscr{P}_p(\mathbb{R}^d)$

(probability measures with finite p-order moment supported on $\mathbb{R}^d$) is defined as

$$W_p^p(\mu,\nu) := \inf\left\{\int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^p d\gamma(x,y) : \ \gamma \in \Pi(\mu,\nu)\right\},$$

where $\Pi(\mu,\nu) := \left\{\gamma \in \mathscr{P}_p(\mathbb{R}^d\times\mathbb{R}^d) : \ (p^x)_\sharp\gamma = \mu, \ (p^y)_\sharp\gamma = \nu\right\}$ stands for the set of admissible transport plans in the transportation of μ onto ν, and $p^x, p^y : \mathbb{R}^d\times\mathbb{R}^d \to \mathbb{R}^d$ denote the canonical projection operators, i.e. $p^x(a,b) = a$ and $p^y(a,b) = b$. We refer to the metric space $(\mathscr{P}_p(\mathbb{R}^d), W_p)$ as the Wasserstein space.

We refer to [AGS08, GT19] and to [CD18a, Chapter 5] for the notion of Wasserstein differentiability and *fully* C^k functions defined on the Wasserstein space, respectively. Based on [Ahu16, CD18a, GMMZ22, MM24] we recall the notion of displacement monotonicity which is given formally in Definition 2.2.1 and Definition 2.2.4.

Definition 5.2.1. Let $G : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ be a fully C^1 function. Based on [GMMZ22, Definition 2.7], we say that G is *displacement semi-monotone* or *displacement α-monotone*, if there exists $\alpha \in \mathbb{R}$ such that $(x,\mu) \mapsto G(x,\mu) + \frac{\alpha}{2}|x|^2$ is displacement monotone.

For the corresponding Hamiltonians, we can define the displacement monotonicity condition as follows.

Definition 5.2.2. Let $H : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ be such that $H(\cdot,\mu,\cdot) \in C^1(\mathbb{R}^d\times\mathbb{R}^d)$ for all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$. We say that H is displacement monotone, if

$$\begin{aligned} &-\int_{\mathbb{R}^d\times\mathbb{R}^d} [\partial_x H(x,\mu,p^1(x)) - \partial_x H(y,\nu,p^2(y))] \cdot (x-y) d\gamma(x,y) \qquad (5.2.1) \\ &+\int_{\mathbb{R}^d\times\mathbb{R}^d} [\partial_p H(x,\mu,p^1(x)) - \partial_p H(y,\nu,p^2(x))] \cdot (p^1(x) - p^2(y)) d\gamma(x,y), \end{aligned}$$

for all $\mu,\nu \in \mathscr{P}_2(\mathbb{R}^d)$, $\gamma \in \Pi(\mu,\nu)$ and for all $p^1, p^2 \in C_b(\mathbb{R}^d;\mathbb{R}^d)$.

Remark 5.2.3. 1. Suppose that $H : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ is fully C^2, strictly convex in the p-variable and satisfies

$$\begin{aligned} &\int_{\mathbb{R}^d\times\mathbb{R}^d} [\partial_{x\mu} H(x,\mu,\tilde{x},p(x))v(\tilde{x}) + \partial_{xx} H(x,\mu,p(x))v(x)] \cdot v(x) d\mu(x) d\mu(\tilde{x}) \qquad (5.2.2) \\ &+ \frac{1}{4}\int_{\mathbb{R}^d} \left\{\left|[\partial_{pp} H(x,\mu,p(x))]^{-\frac{1}{2}} \int_{\mathbb{R}^d} \partial_{p\mu} H(x,\mu,\tilde{x},p(x))v(\tilde{x}) d\mu(\tilde{x})\right|^2\right\} d\mu(x) \le 0, \end{aligned}$$

for all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, for all $p \in C(\mathbb{R}^d;\mathbb{R}^d)$ and for all $v \in L^2_\mu(\mathbb{R}^d;\mathbb{R}^d)$. Then H satisfies the displacement monotonicity condition from Definition 5.2.2. For the proof of this fact we refer to [MM24, Lemma 2.7].

Definition 5.2.4. [MZ24, Definition 3.8],[MZ22b, Definition 3.4] Let $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}^4$ be such that $\lambda_0 > 0$, $\lambda_1 \in \mathbb{R}$, $\lambda_2 > 0$ and $\lambda_3 \geq 0$. Let $G : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ be fully C^2. It is said that G is λ-anti-monotone, if

$$\begin{aligned}
&\lambda_0 \int_{\mathbb{R}^d} \langle \partial_{xx} G(x,\mu)\xi(x), \xi(x)\rangle \, d\mu(x) + \lambda_1 \int_{\mathbb{R}^d\times\mathbb{R}^d} \langle \partial_{x\mu} G(x,\mu,\tilde{x})\xi(x), \xi(\tilde{x})\rangle \, d\mu(x) d\mu(\tilde{x}) \\
&+ \int_{\mathbb{R}^d} \left|\partial_{xx} G(x,\mu)\xi(x)\right|^2 d\mu(x) + \lambda_2 \int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} \partial_{x\mu} G(x,\mu,\tilde{x})\xi(\tilde{x}) \, d\mu(\tilde{x}) \right|^2 d\mu(x) \\
&\leq \lambda_3 \int_{\mathbb{R}^d} |\xi(x)|^2 \, d\mu(x)
\end{aligned}$$

for all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$ and for all $\xi \in L^2_\mu(\mathbb{R}^d;\mathbb{R}^d)$.

5.2.2 New Well-Posedness Theories for MFG and master equations

On the data (H, G) we impose the same assumptions as in Chapter 3, see Assumptions 4 and 5. These are relatively standard assumptions, which appear naturally in the literature on the well-posedness theories for master equations.

We now prove Theorem 5.1.1.

Proof of Theorem 5.1.1. Via direct computation we can verify that V is a solution of the master equation with data (H, G) if and only if $\tilde{V}(t,x,\mu) := V(t,x,\mu) + \frac{\alpha}{2}|x|^2 - \frac{(\beta_0^2+\beta^2)\alpha d}{2}(t - T)$ is a solution of the master equation with data (H_α, G_α). □

Remark 5.2.5. Because of the connection between the solvability of the master equation with data (H, G) and (H_α, G_α) described in Theorem 5.1.1, the same connection holds true for the solutions to the corresponding finite dimensional mean field games systems as well.

Recall the definition (5.1.2). Now we give some sufficient conditions on Hamiltonians H which would result into the displacement monotonicity of the transformed Hamiltonians H_α.

Lemma 5.2.6. *Let H be fully C^2. Then H_α is displacement monotone if and only if*

$$\int_{\mathbb{R}^d} \left[\Big(\partial_{xx} H(x, \mu, p(x)) - 2\alpha \partial_{xp} H(x, \mu, p(x)) \Big) v(x) \right] \cdot v(x) d\mu(x) \tag{5.2.3}$$
$$+ \int_{\mathbb{R}^d \times \mathbb{R}^d} \left[\Big(\partial_{x\mu} H(x, \mu, \tilde{x}, p(x)) - 2\alpha \partial_{p\mu} H(x, \mu, \tilde{x}, p(x)) \Big) v(\tilde{x}) \right] \cdot v(x) d\mu(x) d\mu(\tilde{x})$$
$$+ \frac{1}{4} \int_{\mathbb{R}^d} \left\{ \left| [\partial_{pp} H(x, \mu, p(x))]^{-\frac{1}{2}} \left[\int_{\mathbb{R}^d} \partial_{p\mu} H(x, \mu, \tilde{x}, p(x)) v(\tilde{x}) d\mu(\tilde{x}) \right. \right. \right.$$
$$\left. \left. \left. + 2\alpha \partial_{pp} H(x, \mu, p(x)) v(x) \right] \right|^2 \right\} d\mu(x) \tag{5.2.4}$$
$$\leq 0,$$

for all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, for all $p \in C(\mathbb{R}^d; \mathbb{R}^d)$ and for all $v \in L^2_\mu(\mathbb{R}^d; \mathbb{R}^d)$.

Proof. We readily compute

$$\begin{aligned}
\partial_{xx} \tilde{H}(x, \mu, p) &= \partial_{xx} H(x, \mu, p - \alpha x) - 2\alpha \operatorname{Re}(\partial_{xp} H(x, \mu, p - \alpha x)) + \alpha^2 \partial_{pp} H(x, \mu, p - \alpha x), \\
\partial_{x\mu} \tilde{H}(x, \mu, \cdot, p) &= \partial_{x\mu} H(x, \mu, \cdot, p - \alpha x) - \alpha \partial_{p\mu} H(x, \mu, \cdot, p - \alpha x), \\
\partial_{p\mu} \tilde{H}(x, \mu, \cdot, p) &= \partial_{p\mu} H(x, \mu, \cdot, p - \alpha x), \\
\partial_{pp} \tilde{H}(x, \mu, p) &= \partial_{pp} H(x, \mu, p - \alpha x).
\end{aligned}$$

The result now immediately follows by writing the inequality (5.2.2) for $\tilde{H}$ in terms of H, after noting that we may replace $\operatorname{Re}(\partial_{xp} H)$ with $\partial_{xp} H$ since the quadratic form induced by a skew-symmetric operator is null. □

Remark 5.2.7. The inequality in (5.2.3) can be equivalently rewritten as

$$\int_{\mathbb{R}^d \times \mathbb{R}^d} [\partial_{x\mu} H(x, \mu, \tilde{x}, p(x)) v(\tilde{x}) - \alpha \partial_{p\mu} H(x, \mu, \tilde{x}, p(x)) v(\tilde{x})] \cdot v(x) d\mu(x) d\mu(\tilde{x}) \tag{5.2.5}$$
$$+ \int_{\mathbb{R}^d} \left[\partial_{xx} H(x, \mu, p(x)) v(x) - 2\alpha \partial_{xp} H(x, \mu, p(x)) v(x) + \alpha^2 \partial_{pp} H(x, \mu, p(x)) v(x) \right] \cdot v(x) d\mu(x)$$
$$+ \frac{1}{4} \int_{\mathbb{R}^d} \left\{ \left| [\partial_{pp} H(x, \mu, p(x))]^{-\frac{1}{2}} \int_{\mathbb{R}^d} \partial_{p\mu} H(x, \mu, \tilde{x}, p(x)) v(\tilde{x}) d\mu(\tilde{x}) \right|^2 \right\} d\mu(x) \leq 0,$$

for all $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, for all $p \in C(\mathbb{R}^d; \mathbb{R}^d)$ and for all $v \in L^2_\mu(\mathbb{R}^d; \mathbb{R}^d)$. This is the exact same condition as [MZ22b, (5.10)].

We introduce the following notations.

$$\underline{\kappa}(\partial_{xp}H) := \inf_{(x,\mu,p)\in\mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d)\times\mathbb{R}^d} \lambda_{\min}(\operatorname{Re}\partial_{xp}H(x,\mu,p)),$$

where for $A \in \mathbb{R}^{d\times d}$, we adopt the notation $\operatorname{Re}(A) := (A + A^\top)/2$ and for $A \in \mathbb{R}^{d\times d}$ symmetric $\lambda_{\min}(A)$ stands for its smallest eigenvalue. Furthermore, to denote the suprema of the standard 2-matrix norms, we use the notation

$$|\partial_{x\mu}H| := \sup_{(x,\mu,p,\tilde{x})\in\mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d)\times\mathbb{R}^d\times\mathbb{R}^d} |\partial_{x\mu}H(x,\mu,p,\tilde{x})|\,;$$

$$|\partial_{p\mu}H| := \sup_{(x,\mu,p,\tilde{x})\in\mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d)\times\mathbb{R}^d\times\mathbb{R}^d} |\partial_{p\mu}H(x,\mu,p,\tilde{x})|\,;$$

$$|\partial_{xx}H| := \sup_{(x,\mu,p)\in\mathbb{R}^d\times\mathscr{P}_2(\mathbb{R}^d)\times\mathbb{R}^d} |\partial_{xx}H(x,\mu,p)|\,,$$

and so on for similar quantities. Now, we can formulate the second main result of our chapter.

Theorem 5.2.8. *Suppose that* $H : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ *satisfies* $\partial_{pp}H(x,\mu,p) \geq c_0^{-1}I$, *for some* $c_0 > 0$ *and for all* $(x,\mu,p) \in \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d$. *Suppose that the quantities* $\underline{\kappa}(\partial_{xp}H), |\partial_{pp}H|\,, |\partial_{xx}H|\,, |\partial_{p\mu}H|$ *and* $|\partial_{x\mu}H|$ *are finite. Define*

$$L^H_{our} := |\partial_{x\mu}H| + \frac{1}{4}c_0\,|\partial_{p\mu}H|^2 + |\partial_{xx}H|\,.$$

Suppose that $\underline{\kappa}(\partial_{xp}H) \geq \frac{1}{2}|\partial_{p\mu}H| + \sqrt{|\partial_{pp}H|\,L^H_{our}}$. *Then* H_α *is displacement monotone for any*

$$\alpha \in \left[\alpha^H_-, \alpha^H_+\right],$$

where

$$\alpha^H_\pm := \frac{\underline{\kappa}(\partial_{xp}H) - \frac{1}{2}|\partial_{p\mu}H| \pm \sqrt{\left(\underline{\kappa}(\partial_{xp}H) - \frac{1}{2}|\partial_{p\mu}H|\right)^2 - |\partial_{pp}H|\,L^H_{our}}}{|\partial_{pp}H|}.$$

In particular we have the result for $\alpha := \frac{\underline{\kappa}(\partial_{xp}H) - \frac{1}{2}|\partial_{p\mu}H|}{|\partial_{pp}H|}$.

Proof. For $\alpha \in \left[\alpha_-^H, \alpha_+^H\right]$, $\mu \in \mathscr{P}_2(\mathbb{R}^d)$, $p \in C(\mathbb{R}^d;\mathbb{R}^d)$ and for $v \in L^2_\mu(\mathbb{R}^d;\mathbb{R}^d)$ normalized, i.e. $\int_{\mathbb{R}^d} |v(x)|^2 d\mu = 1$, we compute

$$
\begin{aligned}
&\int_{\mathbb{R}^d\times\mathbb{R}^d} [\partial_{x\mu}H(x,\mu,\tilde{x},p(x))v(\tilde{x}) - \alpha\partial_{p\mu}H(x,\mu,\tilde{x},p(x))v(\tilde{x})]\cdot v(x)d\mu(x)d\mu(\tilde{x})\\
&+\int_{\mathbb{R}^d}\left[\partial_{xx}H(x,\mu,p(x))v(x) - 2\alpha\partial_{xp}H(x,\mu,p(x))v(x) + \alpha^2\partial_{pp}H(x,\mu,p(x))v(x)\right]\cdot v(x)d\mu(x)\\
&+\frac{1}{4}\int_{\mathbb{R}^d}\left\{\left|[\partial_{pp}H(x,\mu,p(x))]^{-\frac{1}{2}}\int_{\mathbb{R}^d}\partial_{p\mu}H(x,\mu,\tilde{x},p(x))v(\tilde{x})d\mu(\tilde{x})\right|^2\right\}d\mu(x)\\
&\le \int_{\mathbb{R}^d\times\mathbb{R}^d}[|\partial_{x\mu}H| + \alpha|\partial_{p\mu}H| + |\partial_{xx}H| - 2\alpha\underline{\kappa}(\partial_{xp}H) + \alpha^2|\partial_{pp}H|]d\mu(x)d\mu(\tilde{x})\\
&+\frac{c_0}{4}\int_{\mathbb{R}^d}\left\{\left|\int_{\mathbb{R}^d}|\partial_{p\mu}H|\,d\mu(\tilde{x})\right|^2\right\}d\mu(x)\\
&= |\partial_{xx}H| - 2\alpha\underline{\kappa}(\partial_{xp}H) + \alpha^2|\partial_{pp}H| + |\partial_{x\mu}H| + \alpha|\partial_{p\mu}H| + \frac{c_0|\partial_{p\mu}H|^2}{4}\\
&= |\partial_{pp}H|\alpha^2 - 2\left(\underline{\kappa}(\partial_{xp}H) - \frac{1}{2}|\partial_{p\mu}H|\right)\alpha + |\partial_{xx}H| + |\partial_{x\mu}H| + \frac{c_0|\partial_{p\mu}H|^2}{4}\\
&= |\partial_{pp}H|\alpha^2 - 2\left(\underline{\kappa}(\partial_{xp}H) - \frac{1}{2}|\partial_{p\mu}H|\right)\alpha + L^H_{our}\\
&\le 0,
\end{aligned}
$$

where in the last inequality we used the sign of the quadratic expression. □

As an immediate consequence of Theorem 5.2.8, we have the well-posedness result in Corollary 5.1.2.

Proof of Corollary 5.1.2. We see that all second order derivatives of $\tilde{H}$ and H match, except the ones involving ∂_{xp}, for which we have

$$\partial_{xp}\tilde{H} = \partial_{xp}H + \alpha I.$$

By the uniform bounds on the corresponding second order derivatives of H, we see that for α sufficiently large, $\tilde{H}$ fulfills the assumptions of Theorem 5.2.8. Increasing α further if necessary, we can ensure that G is displacement α-monotone. Having G displacement α-monotone and H_α displacement monotone would result via Theorem 5.1.1 in the desired global well-posedness result for the master equation. □

5.2.2.1 Our results and previous results on the master equation involving displacement *semi-monotone* data

We notice that the inequality (5.2.3) is precisely the inequality (5.10) from [MZ22b]. This means in particular that [MZ22b, Theorem 5.6] is a direct consequence of Theorem 5.1.1 and Remark 5.2.6 above.

We note that Theorem 5.1.1 shows that we have a global well-posedness theory for the master equation as long as G is displacement semi-monotone and the corresponding $\tilde{H}$ is displacement monotone. In particular, it is enough for these to satisfy the 'first order' monotonicity conditions, in the sense of Definition 5.2.1(1) and (5.2.1). Therefore, Theorem 5.1.1 together with the well-posedness results from [BMM23a] provide a more general result than the one in [MZ22b, Theorem 5.6].

5.2.2.2 Our results and previous results on the master equation involving *anti-monotone* data

Our first objective in this subsection is to show that any function $G : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ which is λ-anti-monotone in the sense of Definition 5.2.4 is actually displacement α-monotone in the sense of Definition 5.2.1(2), where α can be computed explicitly in terms of $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$. We start with some preparatory results.

Remark 5.2.9. G is λ-anti-monotone in the sense of Definition 5.2.4 with $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$ if and only if

$$\int_{\mathbb{R}^d} \left\{ \left| \partial_{xx} G(x,\mu)\xi(x) + \frac{\lambda_0}{2}\xi(x) \right|^2 + \lambda_2 \left| \int_{\mathbb{R}^d} \partial_{x\mu} G(x,\mu,\tilde{x})\xi(\tilde{x}) d\mu(\tilde{x}) + \frac{\lambda_1}{2\lambda_2}\xi(x) \right|^2 \right\} d\mu(x)$$
$$\leq \left(\lambda_3 + \left(\frac{\lambda_0}{2}\right)^2 + \lambda_2 \left(\frac{\lambda_1}{2\lambda_2}\right)^2 \right) \int_{\mathbb{R}^d} |\xi(x)|^2 \, d\mu(x)$$

Proof. This is immediate by an algebraic manipulation after computing the squares. □

Proposition 5.2.10. *If G is λ-anti monotone in the sense of Definition 5.2.4 with $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$, then*

$$\left| \int_{\mathbb{R}^d \times \mathbb{R}^d} \langle \partial_{x\mu} G(x, \mu, \tilde{x})\xi(x), \xi(\tilde{x})\rangle d\mu(x) d\mu(\tilde{x}) \right| \leq \left(\frac{|\lambda_1|}{2\lambda_2} + \sqrt{\frac{\lambda_3}{\lambda_2} + \frac{{\lambda_0}^2}{4\lambda_2} + \left(\frac{\lambda_1}{2\lambda_2}\right)^2} \right) \int_{\mathbb{R}^d} |\xi(x)|^2 \, d\mu(x)$$

and

$$\left| \int_{\mathbb{R}^d} \langle \partial_{xx} G(x, \mu)\xi(x), \xi(x)\rangle d\mu(x) \right| \leq \left(\frac{|\lambda_0|}{2} + \sqrt{\lambda_3 + \left(\frac{\lambda_0}{2}\right)^2 + \lambda_2 \left(\frac{\lambda_1}{2\lambda_2}\right)^2} \right) \int_{\mathbb{R}^d} d\mu(x) \, |\xi(x)|^2$$

In particular G is displacement α_λ-monotone, with

$$\alpha_\lambda \geq \max \left\{ \frac{|\lambda_1|}{2\lambda_2} + \sqrt{\frac{\lambda_3}{\lambda_2} + \frac{{\lambda_0}^2}{4\lambda_2} + \left(\frac{\lambda_1}{2\lambda_2}\right)^2}; \frac{|\lambda_0|}{2} + \sqrt{\lambda_3 + \left(\frac{\lambda_0}{2}\right)^2 + \lambda_2 \left(\frac{\lambda_1}{2\lambda_2}\right)^2} \right\}.$$

Proof. Let us recall that in the definition of λ-anti-monotonicity we have $\lambda_0 > 0$, $\lambda_2 > 0$, $\lambda_3 \geq 0$ and there is no sign restriction on λ_1.

First, let us suppose that $\lambda_1 \neq 0$.

Note that for any $v, w \in \mathbb{R}^d$ and any $C > 0$ we have

$$|\langle v, w\rangle| \leq \frac{C+2}{2} |v|^2 + \frac{1}{2C} |v + w|^2 .$$

With the choice of $v := \frac{\lambda_1}{2\lambda_2}\xi(x)$ and $w := \int_{\mathbb{R}^d} \partial_{x\mu} G(x, \mu, \tilde{x})\xi(\tilde{x}) d\mu(\tilde{x})$, we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^d} \left| \left\langle \int_{\mathbb{R}^d} \partial_{x\mu} G(x, \mu, \tilde{x})\xi(\tilde{x}) d\mu(\tilde{x}), \frac{\lambda_1}{2\lambda_2}\xi(x) \right\rangle \right| d\mu(x) \\
&\leq \int_{\mathbb{R}^d} \left\{ \left(\frac{C}{2} + 1\right) \left| \frac{\lambda_1}{2\lambda_2}\xi(x) \right|^2 + \frac{1}{2C} \left| \int_{\mathbb{R}^d} \partial_{x\mu} G(x, \mu, \tilde{x})\xi(\tilde{x}) d\mu(\tilde{x}) + \frac{\lambda_1}{2\lambda_2}\xi(x) \right|^2 \right\} d\mu(x) \\
&= \int_{\mathbb{R}^d} \Bigg\{ \left(\frac{C}{2} + 1\right) \left| \frac{\lambda_1}{2\lambda_2}\xi(x) \right|^2 \\
&\qquad + \frac{1}{2C\lambda_2} \left(\lambda_2 \left| \int_{\mathbb{R}^d} \partial_{x\mu} G(x, \mu, \tilde{x})\xi(\tilde{x}) d\mu(\tilde{x}) + \frac{\lambda_1}{2\lambda_2}\xi(x) \right|^2 \right) \Bigg\} d\mu(x) \\
&\leq \int_{\mathbb{R}^d} \left\{ \left(\frac{C}{2} + 1\right) \left| \frac{\lambda_1}{2\lambda_2}\xi(x) \right|^2 + \frac{1}{2C\lambda_2} \left(\lambda_3 + \left(\frac{\lambda_0}{2}\right)^2 + \lambda_2 \left(\frac{\lambda_1}{2\lambda_2}\right)^2 \right) |\xi(x)|^2 \right\} d\mu(x)
\end{aligned}$$

where the last inequality follows from Proposition 5.2.9. Hence,

$$\begin{aligned}
&\int_{\mathbb{R}^d} \left| \left\langle \int_{\mathbb{R}^d} \partial_{x\mu} G(x,\mu,\tilde{x})\xi(\tilde{x})\,d\mu(\tilde{x}), \xi(x) \right\rangle \right| d\mu(x) \\
&\leq \left(\left(\frac{C}{2}+1\right)\frac{|\lambda_1|}{2\lambda_2} + \frac{1}{C|\lambda_1|}\left(\lambda_3 + \left(\frac{\lambda_0}{2}\right)^2 + \lambda_2\left(\frac{\lambda_1}{2\lambda_2}\right)^2\right)\right) \int_{\mathbb{R}^d} |\xi(x)|^2\,d\mu(x) \\
&= \left(\frac{|\lambda_1|}{2\lambda_2} + \frac{C|\lambda_1|}{4\lambda_2} + \frac{1}{C|\lambda_1|}\left(\lambda_3 + \left(\frac{\lambda_0}{2}\right)^2 + \lambda_2\left(\frac{\lambda_1}{2\lambda_2}\right)^2\right)\right) \int_{\mathbb{R}^d} |\xi(x)|^2\,d\mu(x).
\end{aligned}$$

We now take $C = \frac{1}{|\lambda_1|}\sqrt{\left(\lambda_3 + \left(\frac{\lambda_0}{2}\right)^2 + \lambda_2\left(\frac{\lambda_1}{2\lambda_2}\right)^2\right)(4\lambda_2)}$ to obtain

$$\begin{aligned}
&\int_{\mathbb{R}^d} \left| \left\langle \int_{\mathbb{R}^d} \partial_{x\mu} G(x,\mu,\tilde{x})\xi(\tilde{x})\,d\mu(\tilde{x}), \xi(x) \right\rangle \right| d\mu(x) \\
&\leq \left(\frac{|\lambda_1|}{2\lambda_2} + 2\sqrt{\frac{\lambda_3 + \left(\frac{\lambda_0}{2}\right)^2 + \lambda_2\left(\frac{\lambda_1}{2\lambda_2}\right)^2}{4\lambda_2}} \right) \int_{\mathbb{R}^d} |\xi(x)|^2\,d\mu(x) \\
&= \left(\frac{|\lambda_1|}{2\lambda_2} + \sqrt{\frac{\lambda_3}{\lambda_2} + \frac{{\lambda_0}^2}{4\lambda_2} + \left(\frac{\lambda_1}{2\lambda_2}\right)^2} \right) \int_{\mathbb{R}^d} |\xi(x)|^2\,d\mu(x)
\end{aligned}$$

Now, as the left hand side of this estimate is continuous at $\lambda_1 = 0$, we can send $\lambda_1 \to 0$, and conclude the claim for general $\lambda_1 \in \mathbb{R}$.

In the same manner with the choice of $v := \frac{\lambda_0}{2}\xi(x)$ and $w := \partial_{xx}G(x,\mu)\xi(x)$, for $C > 0$ arbitrary we get

$$\begin{aligned}
&\int_{\mathbb{R}^d} |\langle \partial_{xx} G(x,\mu)\xi(x), \xi(x)\rangle|\,d\mu(x) \\
&\leq \frac{2}{|\lambda_0|}\int_{\mathbb{R}^d} \left(\frac{C+2}{2}\left|\frac{\lambda_0}{2}\xi(x)\right|^2 + \frac{1}{2C}\left|\partial_{xx}G(x,\mu)\xi(x) + \frac{\lambda_0}{2}\xi(x)\right|^2 \right) d\mu(x) \\
&\leq \frac{2}{|\lambda_0|}\left(\frac{C+2}{2}\left(\frac{\lambda_0^2}{4}\right) + \frac{1}{2C}\left(\lambda_3 + \left(\frac{\lambda_0}{2}\right)^2 + \lambda_2\left(\frac{\lambda_1}{2\lambda_2}\right)^2\right)\right) \int_{\mathbb{R}^d} d\mu(x)\,|\xi(x)|^2 \\
&= \left(\frac{|\lambda_0|}{2} + \frac{C\,|\lambda_0|}{4} + \frac{1}{|\lambda_0|\,C}\left(\lambda_3 + \left(\frac{\lambda_0}{2}\right)^2 + \lambda_2\left(\frac{\lambda_1}{2\lambda_2}\right)^2\right)\right) \int_{\mathbb{R}^d} d\mu(x)\,|\xi(x)|^2
\end{aligned}$$

By taking $C = \frac{2}{|\lambda_0|}\sqrt{\left(\lambda_3 + \left(\frac{\lambda_0}{2}\right)^2 + \lambda_2\left(\frac{\lambda_1}{2\lambda_2}\right)^2\right)}$ we obtain the result. □

Remark 5.2.11. In Proposition 5.2.10 we see that the estimates, and hence the conclusion regarding the displacement α-monotonicity, hold true even for $\lambda_0 \leq 0$. Therefore, we might drop the requirement $\lambda_0 > 0$, and our claims from below will remain true.

Corollary 5.2.12. *Let $G : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ be λ-anti-monotone which satisfies Assumption 4. Suppose that $H : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ satisfies Assumption 5 and it is such that H_{α_λ} is displacement monotone, where the constant α_λ is given in Proposition 5.2.10. Then, the master equation* (1.1) *with data* (H, G) *is globally well-posed.*

Proof. This is a direct consequence of Proposition 5.2.10 and Theorem 5.1.1. □

We would like to conclude our chapter by showing that, if H is strictly convex in the p-variable, then the main theorem on the global well-posedness of the master equation from [MZ24, Theorem 7.1] is a particular case of our main results from Corollary 5.2.12. For completeness, we informally state this here.

Theorem 5.2.13. *[MZ24, Theorem 7.1] Suppose that $G : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d)$ is smooth enough with uniformly bounded second, third and fourth order derivatives. Suppose that the Hamiltonian $H : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ has the specific factorization*

$$H(x, \mu, p) := \langle A_0 x, p \rangle + H_0(x, \mu, p),$$

for a constant matrix $A_0 \in \mathbb{R}^{d \times d}$ and $H_0 : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ smooth enough. Suppose furthermore that G is λ-anti-monotone and that a special set of specific assumption take place jointly for $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$, the matrix A_0 and H_0. Then the master equation (1.1) *is globally well-posed for any $T > 0$, in the classical sense.*

Proposition 5.2.14. *Suppose that $G : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \to \mathbb{R}$ is λ-anti monotone and satisfies Assumption 4. Suppose that $H : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ is given by*

$$H(x, \mu, p) = \langle A_0 x, p \rangle + H_0(x, \mu, p),$$

with $H_0 : \mathbb{R}^d \times \mathscr{P}_2(\mathbb{R}^d) \times \mathbb{R}^d \to \mathbb{R}$ satisfying Assumption 5 and $A_0 \in \mathbb{R}^{d\times d}$ is a given constant matrix. Let $K_H := c_0\,|\partial_{pp}H| = c_0\,|\partial_{pp}H_0|$ be the condition number of $\partial_{pp}H$. Suppose that

$$\underline{\kappa}(A_0) \geq \max\left\{\left(\frac{7}{2} + \frac{\sqrt{K_H}}{2}\right)L_2^{H_0} + \sqrt{|\partial_{pp}H|\,|\partial_{xx}H_0|};\left(\frac{3}{2} + f(\lambda)\right)L_2^{H_0}\right\}, \tag{5.2.6}$$

where $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_2)$, we have set

$$\begin{aligned} f(\lambda) &:= \frac{5|\lambda_1|}{4\lambda_2} + 1 + \frac{\lambda_3}{2\lambda_2} + \frac{\lambda_0}{4\lambda_2} + \frac{5\lambda_0}{4} + \frac{\lambda_3}{2} + \frac{|\lambda_1|}{4} \\ &= 1 + \frac{1}{2}\left(\frac{5\lambda_0}{2} + \frac{|\lambda_1|}{2} + \lambda_3\right) + \frac{1}{2\lambda_2}\left(\frac{\lambda_0}{2} + \frac{5|\lambda_1|}{2} + \lambda_3\right), \end{aligned}$$

and $L_2^{H_0} > 0$ is a constant associated to H_0, satisfying

$$|\partial_{xp}H_0| \leq L_2^{H_0},\ |\partial_{pp}H_0| \leq L_2^{H_0},\ |\partial_{x\mu}H_0| \leq L_2^{H_0} \text{ and } |\partial_{p\mu}H_0| \leq L_2^{H_0}.$$

Then the master equation is globally well-posed.

Proof. Let us note that by the definition of $L_2^{H_0}$ and by the definition of $L_{our}^{H_0}$, we have that

$$L_{our}^{H_0} \leq L_2^{H_0} + \frac{c_0}{4}\left(L^{H_0}\right)^2 + |\partial_{xx}H_0|. \tag{5.2.7}$$

As $\underline{\kappa}(\partial_{xp}H) \geq \underline{\kappa}(A_0) - |\partial_{xp}H_0|$, we see that the assumption $\underline{\kappa}(A_0) \geq (\frac{7}{2} + \frac{\sqrt{K_H}}{2})L_2^{H_0} + \sqrt{|\partial_{pp}H|\,|\partial_{xx}H_0|}$ and (5.2.7) imply

$$\begin{aligned} \underline{\kappa}(\partial_{xp}H) &\geq \underline{\kappa}(A_0) - |\partial_{xp}H_0| \geq 3L_2^{H_0} + \frac{1}{2}\,|\partial_{p\mu}H| - |\partial_{xp}H_0| + \frac{\sqrt{K_H}}{2}L_2^{H_0} + \sqrt{|\partial_{pp}H|\,|\partial_{xx}H_0|} \\ &\geq 2L_2^{H_0} + \frac{1}{2}\,|\partial_{p\mu}H| + \frac{\sqrt{K_H}}{2}L_2^{H_0} + \sqrt{|\partial_{pp}H|\,|\partial_{xx}H_0|} \\ &= \frac{1}{2}\,|\partial_{p\mu}H| + \sqrt{\frac{c_0}{4}|\partial_{pp}H_0|\left(L_2^{H_0}\right)^2} + \sqrt{4\left(L_2^{H_0}\right)^2} + \sqrt{|\partial_{pp}H|\,|\partial_{xx}H_0|} \\ &\geq \frac{1}{2}\,|\partial_{p\mu}H| + \sqrt{\frac{c_0}{4}|\partial_{pp}H_0|\left(L_2^{H_0}\right)^2} + \sqrt{4|\partial_{pp}H_0|L_2^{H_0}} + \sqrt{|\partial_{pp}H|\,|\partial_{xx}H_0|} \\ &\geq \frac{1}{2}\,|\partial_{p\mu}H_0| + \sqrt{|\partial_{pp}H_0|\,L_{our}^{H_0}} \\ &= \frac{1}{2}\,|\partial_{p\mu}H| + \sqrt{|\partial_{pp}H|\,L_{our}^{H}} \end{aligned}$$

and so we can apply Theorem 5.2.8. We get that H is displacement α-monotone with

$$\begin{aligned}\alpha &= \frac{\underline{\kappa}(\partial_{xp}H) - \frac{1}{2}\left|\partial_{p\mu}H\right|}{\left|\partial_{pp}H\right|} \\ &\geq \frac{\underline{\kappa}(A_0) - \left|\partial_{xp}H_0\right| - \frac{1}{2}\left|\partial_{p\mu}H_0\right|}{\left|\partial_{pp}H_0\right|} \\ &\geq \frac{\left(\frac{3}{2} + f(\lambda)\right) L_2^{H_0} - \left|\partial_{xp}H_0\right| - \frac{1}{2}\left|\partial_{p\mu}H_0\right|}{\left|\partial_{pp}H_0\right|} \\ &\geq f(\lambda).\end{aligned}$$

From Proposition 5.2.10 we see that G is semi-monotone with constant

$$\begin{aligned}\eta &:= \frac{|\lambda_1|}{2\lambda_2} + \sqrt{\frac{\lambda_3}{\lambda_2} + \frac{{\lambda_0}^2}{4\lambda_2} + \left(\frac{\lambda_1}{2\lambda_2}\right)^2}^{\,2} + \frac{\lambda_0}{2} + \sqrt{\lambda_3 + \left(\frac{\lambda_0}{2}\right)^2 + \lambda_2\left(\frac{\lambda_1}{2\lambda_2}\right)^2}^{\,2} \\ &\leq \frac{|\lambda_1|}{2\lambda_2} + \sqrt{\frac{\lambda_3}{\lambda_2}} + \sqrt{\frac{{\lambda_0}^2}{4\lambda_2}} + \sqrt{\left(\frac{\lambda_1}{2\lambda_2}\right)^2}^{\,2} + \frac{\lambda_0}{2} + \sqrt{\lambda_3} + \sqrt{\left(\frac{\lambda_0}{2}\right)^2}^{\,2} + \sqrt{\lambda_2\left(\frac{\lambda_1}{2\lambda_2}\right)^2}^{\,2} \\ &\leq \frac{|\lambda_1|}{\lambda_2} + \sqrt{\frac{\lambda_3}{\lambda_2}} + \sqrt{\frac{{\lambda_0}^2}{4\lambda_2}} + \lambda_0 + \sqrt{\lambda_3} + \sqrt{\frac{\lambda_1^2}{4\lambda_2}} \\ &\leq \frac{|\lambda_1|}{\lambda_2} + \frac{1}{2} + \frac{\lambda_3}{2\lambda_2} + \frac{\lambda_0}{4\lambda_2} + \frac{\lambda_0}{4} + \lambda_0 + \frac{1}{2} + \frac{\lambda_3}{2} + \frac{|\lambda_1|}{4\lambda_2} + \frac{|\lambda_1|}{4} \\ &= \frac{5|\lambda_1|}{4\lambda_2} + 1 + \frac{\lambda_3}{2\lambda_2} + \frac{\lambda_0}{4\lambda_2} + \frac{5\lambda_0}{4} + \frac{\lambda_3}{2} + \frac{|\lambda_1|}{4} \\ &= f(\lambda)\end{aligned}$$

and so the result follows. □

Remark 5.2.15. We compare Proposition 5.2.14 with [MZ24, Theorem 7.1]. This theorem has many assumptions. We show that up to constants (depending only on K_H) only a few of these many assumptions imply our assumptions. First, we recall that the definition of the 3×3 matrices A_1, A_2 from formula [MZ24, (4.3)]. These are not constructed from A_0 above, and they involve constants coming in particular from $\lambda = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$. Furthermore, for $A \in \mathbb{R}^{d\times d}$, $\bar{\kappa}(A)$ stands for the largest eigenvalue of $\mathrm{Re}(A)$.

To continue we need the assumption

$$\underline{\kappa}(A_0) \geq (1 + \bar{\kappa}(A_1^{-1}A_2))L_2^{H_0}. \tag{5.2.8}$$

In [MZ24, Theorem 7.1] (specifically the second item of (7.1)) it is assumed that

$$\underline{\kappa}(A_0) \geq (1 + \underline{\kappa}(A_1^{-1}A_2))L_2^{H_0}, \tag{5.2.8$'$}$$

although they probably meant to assume (5.2.8)[1] .

We can formulate the following statement.

Claim. The assumptions of [MZ24, Theorem 7.1], up to a multiplicative constant depending on K_H, imply (5.2.6).

Proof of claim. By definition, we have that $\bar{\kappa}(A_1^{-1}A_2) \geq v^\top A_1^{-1}A_2 v$ for any unit vector $v \in \mathbb{R}^3$. Taking $v = \frac{1}{\sqrt{3}}(1,1,1)^\top$ and using the explicit form of A_1, A_2 given in [MZ24, (4.3)] together with the fact that all the entries of these matrices are non-negative, by direct computation we obtain

$$\begin{aligned}
\bar{\kappa}(A_1^{-1}A_2) &\geq \frac{1}{3}\left(\frac{1}{4}\left(\lambda_0 + \lambda_0 + \left|\lambda_0 - \frac{1}{2}\lambda_1\right| + \lambda_3\right) + \frac{1}{2\lambda_2}\left(\lambda_0 + |\lambda_1| + (\frac{1}{2}|\lambda_1| + \lambda_2 + \lambda_3)\right)\right) \\
&\geq \frac{1}{3}\left(\frac{1}{4}\left(\lambda_0 + \lambda_0 - |\lambda_0| + \frac{1}{2}|\lambda_1| + \lambda_3\right) + \frac{1}{2\lambda_2}\left(\lambda_0 + |\lambda_1| + (\frac{1}{2}|\lambda_1| + \lambda_2 + \lambda_3)\right)\right) \\
&= \frac{1}{3}\left(\frac{1}{4}\left(\lambda_0 + \frac{1}{2}|\lambda_1| + \lambda_3\right) + \frac{1}{2\lambda_2}\left(\lambda_0 + |\lambda_1| + (\frac{1}{2}|\lambda_1| + \lambda_2 + \lambda_3)\right)\right) \\
&\geq \frac{1}{15}f(\lambda),
\end{aligned}$$

so (5.2.8) implies that

$$\underline{\kappa}(A_0) \geq \frac{1}{15}L_2^{H_0}\left(15 + f(\lambda)\right). \tag{5.2.9}$$

Furthermore we see from the second inequality in [MZ24, (7.2)] that $\bar{\gamma}\underline{\kappa}(A_0) \geq |\partial_{xx}H|$. By the assumption (i) of [MZ24, Theorem 7.1] we have that $\bar{\gamma}$ satisfies [MZ24, (4.2)] in which the first inequality implies that $\lambda_0 > \frac{\bar{\gamma}^2}{4\underline{\gamma}} - \frac{8\lambda_3}{4\underline{\gamma}}$. Hence we obtain $(4\underline{\gamma}\lambda_0 + 8\lambda_3) \geq \bar{\gamma}^2$. It is

[1] The $\underline{\kappa}$ on the right-hand side is likely a typo as in the fourth to last line on [MZ24, page 15] the authors need to use $\bar{\kappa}(A_1^{-1}A_2)$. Furthermore we see $\bar{\kappa}$ appearing correctly also in a similar assumption, [MZ22b, (6.3)].

clear that $2f(\lambda) \geq \lambda_0$ and $2f(\lambda) \geq \lambda_3$, therefore we get $16f(\lambda)(1+\underline{\gamma}) \geq \bar{\gamma}^2$. Since $\underline{\gamma} < \bar{\gamma}$ by assumption (i) of [MZ24, Theorem 7.1] and $1 < \bar{\gamma}$ by the same assumption we get $2\bar{\gamma} \geq 1+\underline{\gamma}$ and so we obtain $32f(\lambda) \geq \bar{\gamma}$. Hence we get

$$\underline{\kappa}(A_0)^2 \geq \frac{L_2^{H_0}}{15} f(\lambda)\underline{\kappa}(A_0) \geq \frac{L_2^{H_0}}{15 \cdot 32}\bar{\gamma}\underline{\kappa}(A_0) \geq \frac{|\partial_{pp}H|}{15 \cdot 32}|\partial_{xx}H|$$

and so we obtain $\underline{\kappa}(A_0) \geq \frac{1}{4\sqrt{30}}\sqrt{|\partial_{pp}H|\,|\partial_{xx}H|}$.

Moreover, (5.2.8) implies that $\underline{\kappa}(A_0) \geq L_2^{H_0}$ and so we get

$$\underline{\kappa}(A_0) \geq \frac{1}{2}L_2^{H_0} + \frac{1}{8\sqrt{30}}\sqrt{|\partial_{pp}H|\,|\partial_{xx}H|}. \tag{5.2.10}$$

To summarize, the assumptions of [MZ24, Theorem 7.1] imply (5.2.9) and (5.2.10) which in turn imply that

$$\underline{\kappa}(A_0) \geq \frac{1}{8\sqrt{30}+\sqrt{K_H}} \max\left\{\left(\frac{7}{2}+\frac{\sqrt{K_H}}{2}\right)L_2^{H_0} + \sqrt{|\partial_{pp}H|\,|\partial_{xx}H_0|}; \left(\frac{3}{2}+f(\lambda)\right)L_2^{H_0}\right\}.$$

This, aside from the constant of $\frac{1}{8\sqrt{30}+\sqrt{K_H}}$ in front, is the exact assumption (5.2.6) of our Proposition 5.2.14.

CHAPTER 6

Optimal Transport with Quantum Dynamics

6.1 Introduction

The classical optimal transport (OT) problem, introduced by Gaspard Monge in the 18th century and further developed by Leonid Kantorovich in the 20th century, is a foundational question in both pure and applied mathematics. The problem focuses on finding the most efficient way to transport one distribution of mass, described by a probability measure, to another, under a specified cost function. In this chapter, we explore an extension of this problem to quantum mechanics, where the transport is governed by quantum dynamics, specifically the Schrödinger equation. We show how this quantum dynamic extension of OT is connected to a well-known problem in quantum mechanics: the Pauli problem.

6.1.1 The Classical Optimal Transport Problem

6.1.1.1 Formulation of the Classical OT Problem

The classical optimal transport problem can be formulated as follows: given two probability measures μ_0 and μ_1 on $\mathbb{R}^d$, the goal is to find a transport map $T : \mathbb{R}^d \to \mathbb{R}^d$ such that T pushes forward μ_0 to μ_1, i.e., $T_{\#}\mu_0 = \mu_1$, while minimizing a given cost functional. The cost of transporting a unit of mass from x to $T(x)$ is typically described by the Euclidean distance $|x - T(x)|$, or more generally by a cost function $c(x, T(x))$.

The problem is formally stated as:

$$\inf_T \int_{\mathbb{R}^d} c(x, T(x))\, d\mu_0(x)$$

subject to the constraint $T_{\#}\mu_0 = \mu_1$, i.e., for all measurable sets $A \subset \mathbb{R}^d$, we have

$$\mu_1(A) = \mu_0(T^{-1}(A)).$$

While originally posed by Monge, the problem proved challenging due to its nonlinearity. Kantorovich's relaxation of the problem, formulated in terms of couplings between μ_0 and μ_1, led to significant breakthroughs.

6.1.1.2 Motivation for the Classical OT Problem

The classical OT problem has widespread applications across fields such as logistics, economics, image processing, and machine learning. Initially, it was intended to solve practical problems like minimizing the cost of transporting soil for construction, but its mathematical structure has allowed it to be applied to a variety of domains. In economics, for example, OT can model the most efficient way to allocate resources, while in machine learning, it plays a role in domain adaptation and generative modeling, where transforming one data distribution into another is a key challenge.

Additionally, the OT problem is deeply connected to areas like partial differential equations, geometry, and probability theory. The development of Wasserstein distances, a family of metrics used to measure the "cost" of transforming one probability distribution into another, has led to significant advances in probability theory on metric spaces and has influenced the study of the geometry of probability measures.

6.1.2 Optimal Transport with Quantum Dynamics

6.1.2.1 Formulation of Optimal Transport with Quantum Dynamics

In the classical OT problem, the transport of mass is governed by a deterministic map T. In the quantum dynamic extension of OT, however, the transport of probability densities is governed by the evolution of a wave function under the Schrödinger equation. The goal is to minimize the kinetic energy of a quantum system while subject to constraints on the initial and final probability densities.

Given two probability measures μ_0 and μ_1, we seek to minimize

$$\inf_{\psi_t} \int_0^1 \int_{\mathbb{R}^d} |\nabla \psi_t(x)|^2 \, dx \, dt$$

over all wave functions ψ_t such that $|\psi_0|^2 \, dx = \mu_0$, $|\psi_1|^2 \, dx = \mu_1$, and where ψ_t evolves according to the Schrödinger equation

$$i\hbar \partial_t \psi_t = -\frac{\hbar^2}{2} \Delta \psi_t,$$

with $\hbar = \frac{1}{2\pi}$ being the reduced Planck constant. Here, the cost functional represents the kinetic energy of the quantum system, and the Schrödinger equation governs the wave function's time evolution.

This problem was first considered by Eric Carlin and Wilfrid Gangbo (private communication).

6.1.2.2 Motivation for Optimal Transport with Quantum Dynamics

There are several compelling reasons to explore optimal transport with quantum dynamics. First, quantum mechanics is a fundamental theory in physics, and the Schrödinger equation describes the evolution of quantum systems. Extending OT into this quantum dynamic framework opens up possibilities for studying the interaction between quantum systems and probability distributions.

Second, this extension provides a natural bridge between classical mechanics and quantum mechanics. As the reduced Planck constant $\hbar \to 0$, the quantum dynamic version of OT reduces to the classical OT problem, linking both regimes under one conceptual framework. This connection offers new insights into transport problems in both classical and quantum settings, and allows for the application of methods from both fields.

6.1.3 The Pauli Problem

6.1.3.1 Formulation of the Pauli Problem

The Pauli problem, named after the physicist Wolfgang Pauli, asks whether a quantum state can be uniquely determined from the knowledge of its magnitude in both position and momentum space (via its Fourier transform). More formally, given two probability measures μ_0 and μ_1 on $\mathbb{R}^d$, we define the set

$$S_1 = \left\{ \psi \in L^2 : |\psi|^2 \, dx = \mu_0, \, \left|\hat{\psi}\right|^2 d\xi = \mu_1 \right\},$$

where $\hat{\psi}$ denotes the Fourier transform of ψ. The Pauli problem asks whether S_1 is a singleton (up to multiplication by a complex phase) or whether multiple quantum states could correspond to the same magnitude distributions in both physical and momentum space.

This problem is deeply related to the uncertainty principle, which places limits on the simultaneous knowledge of a quantum particle's position and momentum. In this sense, the Pauli problem can be interpreted as a question about the uniqueness of the mapping between a quantum state and its probability distributions in position and momentum space.

6.1.3.2 Motivation for the Pauli Problem

The Pauli problem addresses fundamental questions in quantum mechanics. In practical quantum measurements, information is often obtained about the magnitude of the wave function in position or momentum space, but not its full complex phase. The Pauli problem asks whether this limited information is sufficient to uniquely determine the quantum state, which has important implications for fields like quantum state tomography and quantum information theory.

Furthermore, the Pauli problem highlights the interplay between position and momentum, which are conjugate variables in quantum mechanics. Understanding the conditions under which these magnitude distributions in position and momentum space determine a unique quantum state is crucial for exploring the boundaries of quantum measurements and the uncertainty principle.

6.1.4 Results

We demonstrate that, the optimal transport problem with quantum dynamics always admits a minimizer, provided that the feasible set is non-empty. This ensures that for any given pair of initial and final probability measures, there exists a wave function evolution, governed by quantum dynamics, that optimally transports the probability distribution in a way that minimizes the associated cost functional.

We further explore the relationship between the optimal transport problem with quantum dynamics and the Pauli problem. In particular, we show that the optimal transport problem with quantum dynamics is equivalent to finding a unique (up to symmetry) solution in the feasible set. This means that solving the quantum optimal transport problem can be seen as a means to uniquely determine the quantum state, constrained by probability densities in both physical and Fourier space, similar to the Pauli problem. The symmetry inherent in quantum systems, where states may be determined only up to a phase factor, plays a crucial role in this equivalence.

The Pauli problem, which asks whether a quantum state can be reconstructed from its probability distributions in position and momentum (Fourier) space, can thus be reframed as an instance of an optimal transport problem in the quantum setting. By understanding the structure of the feasible set and developing solutions to the optimal transport problem with quantum dynamics, we provide a new approach to understand the Pauli problem. Specifically, this framework allows us to select a distinguished solution to the Pauli problem from among the possible quantum states that satisfy the given constraints.

This equivalence between the Pauli problem and the optimal transport problem with quantum dynamics is particularly valuable because it brings together two seemingly distinct areas of mathematics and physics.

Moreover, the connection we establish between these two problems opens up new avenues for research. On one hand, it deepens our understanding of quantum transport processes by providing a bridge between classical transport theory and quantum mechanics. On the other hand, it offers new methodologies for tackling problems in quantum state reconstruction, such as those encountered in quantum information theory, quantum state tomography, and even in foundational questions related to the uncertainty principle.

6.2 Optimal Transport with Quantum Dynamics

6.2.1 Problem Description

Say that we are given probability measures μ_0, μ_1.

We consider

$$\inf_{\psi_t} \int_0^1 \int_{\mathbb{R}^d} |\nabla \psi_t(x)|^2 \, dx dt$$

where the infimum is taken over all ψ_t such that $|\psi_0|^2 \, dx = \mu_0$ and $|\psi_1|^2 \, dx = \mu_1$ and ψ_t solves the Schrödinger equation $i\hbar \partial_t \psi_t = -\frac{\hbar^2}{2} \Delta \psi_t$. Here $\hbar = \frac{1}{2\pi}$ is the reduced Planck constant (we take $h = 1$).

Of course, it is easy to see that for general μ_0, μ_1 there might not even exist a single ψ_t that satisfies our constraints. Hence, we assume that μ_0, μ_1 are such that there exists at least one ψ_t that satisfies the constraints.

6.2.2 Some Basic Facts about the Schrödinger Equation

In order that we can be rigorous we will assume that μ_0, μ_1 are absolutely continuous with respect to Lebesgue measure.

If ψ_0 is 'nice' (say Schwartz for example) then we may solve the Schrödinger equation explicitly. On the Fourier side the solution is

$$\hat{\psi}_t(\xi) = e^{-2\pi i(|\xi|^2/2)t}\hat{\psi}_0(\xi) \tag{6.2.1}$$

See [Hal13, Proposition 4.3] for a proof.

Note that this expression makes sense and defines a unique $\psi_t \in L^2$ even if ψ_0 isn't 'nice'. All we need is $\psi_0 \in L^2$ which is guaranteed by $|\psi_0|^2\, dx$ being a probability measure. Hence we will take (6.2.1) as our notion of weak solution of the Schrödinger equation even when ψ_0 isn't nice.

Lemma 6.2.1. *Suppose that ψ_t solves the Schrödinger equation. Then $\int_{\mathbb{R}^d} |\nabla\psi_t(x)|^2\, dx$ is constant in t.*

Proof. By Pancherel's theorem, we have that

$$\int_{\mathbb{R}^d} |\nabla\psi_t(x)|^2\, dx = \int_{\mathbb{R}^d} \left|\xi\hat{\psi}_t(\xi)\right|^2 d\xi = \int_{\mathbb{R}^d} \left|\xi\hat{\psi}_0(\xi)\right|^2 d\xi = \int_{\mathbb{R}^d} |\nabla\psi_0(x)|^2\, dx$$

□

Remark 6.2.2. The above lemma is obvious in physics. Since we have taken the free particle Schrödinger equation, there aren't any forces acting on the particle and hence the kinetic energy is conserved.

6.2.3 Existence of a Minimizer

By Lemma 6.2.1, we have that

$$\int_0^1 \int_{\mathbb{R}^d} |\nabla \psi_t(x)|^2 \, dxdt = \int_{\mathbb{R}^d} |\nabla \psi_0(x)|^2 \, dx$$

and so our problem is reduced to

$$\inf_{\psi_0} \int_{\mathbb{R}^d} |\nabla \psi_0(x)|^2 \, dx$$

where the infimum is taken over all ψ_0 such that $|\psi_0|^2 \, dx = \mu_0$ and $|\psi_1|^2 \, dx = \mu_1$ where $\hat{\psi}_1 = e^{-2\pi i(|\xi|^2/2)} \hat{\psi}_0(\xi)$ is the solution to the Schrödinger equation.

Proposition 6.2.3. *The above problem has a minimizer.*

Proof. Let ψ_0^n be a minimizing sequence. Then

$$\int_{\mathbb{R}^d} |\nabla \psi_0^n(x)|^2 \, dx \leq \int_{\mathbb{R}^d} \left|\nabla \psi_0^0(x)\right|^2 dx$$

and so ψ_0^n is uniformly bounded in H^1. Hence there exists a ψ_0^* and subsequence of ψ_0^n (which we do not relabel) such that $\psi_0^n \to \psi_0^*$ in L^2 and $\psi_0^n \rightharpoonup \psi_0^*$ in H^1.

We claim that ψ_0^* is a minimizer. Indeed since the functional that we seek to minimize is weakly lower semi-continuous we have that

$$\int |\nabla \psi_0^*(x)|^2 \, dx \leq \liminf_{n \to \infty} \int |\nabla \psi_0^n(x)|^2 \, dx = \inf_{\psi_0} \int |\nabla \psi_0(x)|^2 \, dx$$

Furthermore since $\psi_0^n \to \psi_0^*$ in L^2 we get that $|\psi_0^*|^2 = \lim_{n\to\infty} |\psi_0^n|^2 = \mu_0$.

Since (in L^2) $\psi_0^n \to \psi_0^*$ we get that $\hat{\psi}_0^n \to \hat{\psi}_0^*$. From (6.2.1) we get that $\hat{\psi}_1^n \to \hat{\psi}_1^*$ and so $\psi_1^n \to \psi_1^*$ in L^2. Hence we also get that $|\psi_1^*|^2 = \lim_{n\to\infty} |\psi_1^n|^2 = \mu_1$.

Hence ψ_0^* satisfies the constraints and is a minimizer. □

6.2.4 Remarks on the Constraint Set

Let us use $L^2 = L^2(\mathbb{R}^d, \mathbb{C})$ to denote the space of square integrable functions on $\mathbb{R}^d$. We define the unitary operator $P = e^{i\hbar\Delta/2}$ which is the solution operator to the Schrödinger equation.

We saw that the problem is reduced to

$$\inf_{\psi_0} \int_{\mathbb{R}^d} |\nabla \psi_0(x)|^2 \, dx$$

where the infimum is taken over all ψ_0 such that $|\psi_0|^2 \, dx = \mu_0$ and $|P\psi_0|^2 \, dx = \mu_1$. We define $S_0(\mu_0, \mu_1) = \{\psi \in L^2 : |\psi|^2 \, dx = \mu_0, |P\psi|^2 \, dx = \mu_1\}$ which is the set that we are minimizing over.

Fix $\alpha \in \mathbb{R}$. Note that if $\psi \in S_0(\mu_0, \mu_1)$ then $e^{i\alpha}\psi \in S_0(\mu_0, \mu_1)$. We wonder when S_0 is actually a singleton up to this symmetry, i.e. for which μ_0, μ_1 does there exist some $\psi \in L^2$ so that $S_0(\mu_0, \mu_1) = \{e^{i\alpha}\psi : \alpha \in \mathbb{R}\}$.

6.3 The Pauli Problem

6.3.1 Problem Description

The physicist Wolfgang Pauli conjectured whether or not a function in L^2 can be recovered from its magnitude and the magnitude of its Fourier transform. More precisely, given measures μ_0, μ_1 on $\mathbb{R}^d$ define $S_1 = \{\psi \in L^2 : |\psi|^2 \, dx = \mu_0, \left|\hat{\psi}\right|^2 d\xi = \mu_1\}$. The Pauli problem asks whether S_1 is a singleton up to the symmetry $\psi \mapsto e^{i\alpha}\psi$ for some $\alpha \in \mathbb{R}$.

First it is easy to see that S_1 could be empty. For example the Heisenberg uncertainty principle tells us that if μ_0, μ_1 are both points masses then S_1 is empty. Even more, it tells us that if S_1 is non-empty then the product of the variances of μ_0, μ_1 is bounded below by a universal constant.

However even excluding the case where S_1 is empty the conjecture is not true. One can construct specific examples where S_1 is not a singleton (up to symmetry).

The question then becomes to understand when S_1 is a singleton (up to symmetry). A few classes of measures are known but it seems very little is known.

6.3.2 Reformulation of the Problem

We use the notation $p_j = -i\hbar\frac{\partial}{\partial x_j}$ as a (densely defined) operator on L^2. Similarly we use x_j as the operator of multiplication by x_j.

If we have a list of operators $A^1, \ldots, A^d$ on L^2, given a continuous function $f : \mathbb{R}^d \to \mathbb{C}$ we use the notation $f(A) = f(A^1, \ldots, A^d)$ which is defined by the continuous functional calculus (considering L^2 as a C^* algebra for example). We note that $f(A)$ is a (densely defined) operator on L^2.

Lemma 6.3.1. *Fix $\psi \in L^2$. Let U be any unitary operator on L^2 and μ be any measure on $\mathbb{R}^d$. Define $A^j = U^* x_j U$.*

Then $|U\psi|^2 \, dx = \mu$ if and only if $\langle f(A)\psi, \psi\rangle = \int f(y) d\mu(y)$ for all continuous and compactly supported functions f.

Proof. We have

$$\langle f(A)\psi, \psi\rangle = \langle U^* f(A) U\psi, \psi\rangle = \langle f(A)U\psi, U\psi\rangle = \int f(y) \, |U\psi(y)|^2 \, dy$$

and so the result follows. □

Note that if we take U to be the Fourier transform operator then $A^j = p_j$ and so the above lemma tells us that $\left|\hat{\psi}\right|^2 d\xi = \mu$ if and only if $\langle f(p)\psi, \psi\rangle = \int f(y) d\mu(y)$ for all continuous and compactly supported functions f. Hence we have

Corollary 6.3.2.

$$S_1(\mu_0, \mu_1) = \left\{ \phi \in L^2 : \langle \phi, f(x)\phi \rangle = \int f(y) d\mu_0(y), \langle \phi, g(p)\phi \rangle = \int g(y) d\mu_0(y) \right.$$
$$\left. \textit{for all } f, g \textit{ continuous and compact} \right\}$$

6.4 Equivalence of the Two Problems

In this section we show that the two problems are equivalent. In particular we show that S_0 is isometric to S_1.

We start with a quick lemma about commutators which I think is well-known in the physics.

Lemma 6.4.1. *Let A, B be operators such that $[A, [A, B]] = 0$. Then if f is a smooth function we have $[f(A), B] = [A, B]f'(A)$.*

Proof. If $f(x) = x^n$ then the result is clear. Both sides are linear in f so it extends to polynomials and then to smooth functions by density. □

We define the unitary operator

$$P = e^{-i\Delta/2}$$

which is the solution operator to the Schrödinger equation. Also define the operators $K^j = P^* x_j P$.

Lemma 6.4.2. $K^j = x_j + p_j$.

Proof. Note that by the product rule for Laplacian we have $[\frac{-i\hbar}{2}\Delta, x_j] = p_j$. In particular $[\frac{-i\hbar}{2}\Delta, [\frac{-i\hbar}{2}\Delta, x_j]] = 0$ since p_j commutes with Δ. Hence by the previous lemma we have that $[P^*, x_j] = [e^{-i\hbar\Delta/2}, x_j] = p_j e^{-i\hbar\Delta/2} = p_j P^*$. The result now follows by multiplying by P on the right. □

Remark 6.4.3. To a physicist, this is just Newton's first law of motion.

We recall $S_0(\mu_0, \mu_1) = \{\psi \in L^2 : |\psi|^2 \, dx = \mu_0, |P\psi|^2 \, dx = \mu_1\}$. By taking $U = P$ in Lemma 6.3.1 we see that $S_0(\mu_0, \mu_1) = \{\psi \in L^2 : \langle f(x)\psi, \psi\rangle = \int f(y) d\mu_0(y), \langle g(K)\psi, \psi\rangle = \int g(y) d\mu_1(y)$ for all $f, g\}$.

Let Q be the unitary operator on L^2 given by $Q\psi = e^{\frac{i}{\hbar}|x|^2/2}\psi$. We see Q is left multiplication by $e^{\frac{i}{\hbar}|x|^2/2}$ and by the standard abuse of notation we write $Q = e^{\frac{i}{\hbar}|x|^2/2}$.

Proposition 6.4.4. *Fix μ_0, μ_1. Then $S_1(\mu_0, \mu_1) = \{Q\psi : \psi \in S_0(\mu_0, \mu_1)\}$.*

Proof. Fix $\psi \in S_0(\mu_0, \mu_1)$ and let $\phi = Q\psi$. Clearly $|\phi|^2 = |\psi|^2$ and so $\langle \psi, f(A_0)\psi\rangle = \int f(x) d\mu_0(x)$ if and only if $\langle \phi, f(A_0)\phi\rangle = \int f(x) d\mu_0(x)$.

We have $p_j Q = p_j e^{\frac{i}{\hbar}|x|^2/2} = e^{\frac{i}{\hbar}|x|^2/2}(x_j + p_j) = QK^j$ and so $Q^* p_j Q = K^j$. Hence for any g continuous we have $Q^* g(p) Q = g(K)$. This gives

$$\langle g(p)\phi, \phi\rangle = \langle g(p)Q\psi, Q\psi\rangle = \langle Q^* g(p) Q\psi, \psi\rangle = \langle g(K)\psi, \psi\rangle$$

and so $\langle g(K)\psi, \psi\rangle = \int g(x) d\mu_1(x)$ if and only if $\langle g(p)\phi, \phi\rangle = \int g(x) d\mu_1(x)$. □

Remark 6.4.5. The above Proposition says that S_0 is isometric to S_1 via the isometry on L^2 given by Q.

Remark 6.4.6. This isometry didn't come out of nowhere. It represents the canonical transformation that that was studied in Chapter 4. Essentially this transforms the free particular problem (Optimal Transport with Quantum Dynamics) into a simple harmonic oscillator problem (Pauli problem).

6.5 Revisiting Bohm's Interpretation of Quantum Mechanics

Let us return to our original formulation of the Optimal Transport problem with Quantum Dynamics. Given μ_0, μ_1 we want to study the Schrödinger equation: $i\partial_t \psi_t = -\frac{\hbar}{2}\Delta\psi_t$ subject to the boundary conditions $|\psi_0|^2 \, dx = \mu_0$ and $|\psi_1|^2 \, dx = \mu_1$.

In the early 1950s David Bohm proposed writing the Schrödinger equation in polar coordinates. Let $\psi_t = \sqrt{\rho_t} e^{i\theta_t}$ and then the Schrödinger equation becomes

$$\partial_t S + \frac{1}{2} |\nabla S|^2 - \frac{\hbar^2}{2\sqrt{\rho}} \Delta(\sqrt{\rho}) = 0$$

$$\partial_t \rho + \nabla \cdot (\rho \nabla S) = 0$$

and our boundary conditions become $\rho_0 = \mu_0$ and $\rho_1 = \mu_1$. This is now a mean field game system with the Hamilton-Jacobi equation and the continuity equation. The Hamiltonian is $H(x, \mu, p) = \frac{1}{2} |p|^2 - \frac{\hbar^2}{2\sqrt{\mu(x)}} \Delta(\sqrt{\mu(x)})$.

For the rest of this section we will proceed formally.

Let us define the energy $E(\rho) = \frac{1}{2} \int \left| \nabla \sqrt{\rho(x)} \right|^2 dx$. We compute its first variation

$$\begin{aligned} E(\rho + \epsilon) &= \frac{1}{2} \int \left| \nabla \sqrt{\rho(x) + \epsilon(x)} \right|^2 dx \\ &= \frac{1}{2} \int \left| \nabla \left(\sqrt{\rho(x)} + \frac{\epsilon(x)}{2\sqrt{\rho(x)}} \right) \right|^2 dx + o(\epsilon) \\ &= \frac{1}{2} \int \left| \nabla \sqrt{\rho(x)} + \nabla \frac{\epsilon(x)}{2\sqrt{\rho(x)}} \right|^2 dx + o(\epsilon) \\ &= E(\rho) + 1 \int \nabla \sqrt{\rho(x)} \cdot \nabla \frac{\epsilon(x)}{2\sqrt{\rho(x)}} dx + o(\epsilon) \\ &= E(\rho) - \frac{1}{2} \int \frac{\Delta \sqrt{\rho(x)}}{\sqrt{\rho(x)}} \epsilon(x) dx + o(\epsilon) \end{aligned}$$

and so we get

$$\frac{\delta E}{\delta \rho} = -\frac{1}{2\sqrt{\rho(x)}} \Delta \sqrt{\rho(x)}$$

Hence if we define $f(x, \mu) = -\frac{1}{2\sqrt{\rho(x)}} \Delta \sqrt{\rho(x)}$ we get $\partial_\mu E = \nabla f$. Hence our mean field game was actually a potential mean field game with Hamiltonian

$$\mathcal{H}(\mu, b) = \frac{1}{2} \int |b(\omega)|^2 d\omega + \hbar^2 E(\mu)$$

and associated Hamilton Jacobi Equation

$$\partial_t U(t,\mu) + \frac{1}{2}\int |\partial_\mu U(t,\mu,\tilde{x})|^2 \, d\mu(\tilde{x}) + \hbar^2 E(\mu) = 0$$
$$U(0,\mu) = \delta_{\mu_0}(\mu)$$

Because this Hamiltonian is convex in b we may write down an equivalent control problem. We want to minimize

$$\min_{\rho_t, v_t} \int_0^1 \int |v_t(x)|^2 \rho_t(x) dx dt - \hbar^2 \int_0^1 E(\rho_t) dt$$

subject to the continuity equation $\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0$ and the boundary conditions $\rho_0 = \mu_0$ and $\rho_1 = \mu_1$. We remark that as $\hbar \to 0$ we get the classical optimal transport problem. Landau comments that it is important to note that this approximation is accurate not just to $\hbar$ but to $\hbar^2$.

Remark 6.5.1. We have

$$\begin{aligned} E(\rho) &= \frac{1}{2}\int \left|\nabla \sqrt{\rho(x)}\right|^2 dx \\ &= \frac{1}{2}\int \left|\frac{\nabla \sqrt{\rho(x)}}{\sqrt{\rho(x)}}\right|^2 \rho(x) dx \\ &= \frac{1}{2}\int \left|\nabla \ln(\sqrt{\rho(x)})\right|^2 \rho(x) dx \\ &= \frac{1}{8}\int |\nabla \ln(\rho(x))|^2 \rho(x) dx \end{aligned}$$

This is the Fisher information.

REFERENCES

[AC99] P. Albano and P. Cannarsa. Structural properties of singularities of semiconcave functions. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 28(4):719–740, 1999.

[AGS08] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[Ahu16] S. Ahuja. Wellposedness of mean field games with common noise under a weak monotonicity condition. *SIAM J. Control Optim.*, 54(1):30–48, 2016.

[AM23] D.M. Ambrose and A.R. Mészáros. Well-posedness of mean field games master equations involving non-separable local hamiltonians. *Trans. Amer. Math. Soc.*, 376(4):2481–2523, 2023.

[Arn89] V. I. Arnol′d. *Mathematical methods of classical mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1989. Translated from the Russian by K. Vogtmann and A. Weinstein.

[ARY19] S. Ahuja, W. Ren, and T.-W. Yang. Forward-backward stochastic differential equations with monotone functionals and mean field games with common noise. *Stochastic Process. Appl.*, 129(10):3859–3892, 2019.

[BCQ10] R. Buckdahn, P. Cannarsa, and M. Quincampoix. Lipschitz continuity and semiconcavity properties of the value function of a stochastic control problem. *NoDEA Nonlinear Differential Equations Appl.*, 17(6):715–728, 2010.

[BE84] M. Bardi and L. C. Evans. On Hopf's formulas for solutions of Hamilton-Jacobi equations. *Nonlinear Anal.*, 8(11):1373–1381, 1984.

[Ber21] C. Bertucci. Monotone solutions for mean field games master equations: finite state space and optimal stopping. *J. Éc. polytech. Math.*, 8:1099–1132, 2021.

[BGY24] A. Bensoussan, P.J. Graber, and S.C.P. Yam. Control on Hilbert spaces and application to some mean field type control problems. *Ann. Appl. Probab., to appear*, arXiv:2005.10770, 2024+.

[BM24a] M. Bansil and A.R. Mészáros. On classical solutions and canonical transformations for Hamilton–Jacobi–Bellman equations. *preprint*, 2024.

[BM24b] Mohit Bansil and Alpár R. Mészáros. Hidden monotonicity and canonical transformations for mean field games and master equations, 2024.

[BM24c] Mohit Bansil and Alpár R. Mészáros. On classical solutions and canonical transformations for hamilton–jacobi–bellman equations, 2024.

[BMM23a] M. Bansil, A.R. Mészáros, and C. Mou. Global well-posedness of displacement monotone degenerate mean field games master equations. *arXiv:2308.16167*, 2023.

[BMM23b] Mohit Bansil, Alpár R. Mészáros, and Chenchen Mou. Global well-posedness of displacement monotone degenerate mean field games master equations, 2023.

[CC21] P. Cannarsa and W. Cheng. Singularities of solutions of Hamilton-Jacobi equations. *Milan J. Math.*, 89(1):187–215, 2021.

[CCD22] J.-F. Chassagneux, D. Crisan, and F. Delarue. A probabilistic approach to classical solutions of the master equation for large population equilibria. *Mem. Amer. Math. Soc.*, 280(1379):v+123, 2022.

[CCP23] P. Cardaliaguet, M. Cirant, and A. Porretta. Splitting methods and short time existence for the master equations in mean field games. *J. Eur. Math. Soc. (JEMS)*, 25(5):1823–1918, 2023.

[CD15] R. Carmona and F. Delarue. Forward-backward stochastic differential equations and controlled McKean-Vlasov dynamics. *Ann. Probab.*, 43(5):2647–2700, 2015.

[CD18a] R. Carmona and F. Delarue. *Probabilistic theory of mean field games with applications. I. Mean field FBSDEs, control, and games*, volume 83 of *Probability Theory and Stochastic Modelling*. Springer, Cham, 2018.

[CD18b] R. Carmona and F. Delarue. *Probabilistic theory of mean field games with applications II. Mean field games with common noise and master equations*, volume 84 of *Probab. Theory Stoch. Model.* Cham: Springer, 2018.

[CD24] A. Cecchin and F. Delarue. Weak solutions to the master equation of potential mean field games. *Mem. Amer. Math. Soc., to appear*, arXiv:2204.04315, 2024+.

[CDL16] R. Carmona, F. Delarue, and D. Lacker. Mean field games with common noise. *Ann. Probab.*, 44(6):3740–3803, 2016.

[CDLL19] P. Cardaliaguet, F. Delarue, J.-M. Lasry, and P.-L. Lions. *The master equation and the convergence problem in mean field games*, volume 201 of *Annals of Mathematics Studies*. Princeton University Press, Princeton, NJ, 2019.

[CF91] P. Cannarsa and H. Frankowska. Some characterizations of optimal trajectories in control theory. *SIAM J. Control Optim.*, 29(6):1322–1347, 1991.

[CF14] P. Cannarsa and H. Frankowska. From pointwise to local regularity for solutions of Hamilton-Jacobi equations. *Calc. Var. Partial Differential Equations*, 49(3-4):1061–1074, 2014.

[CH78] J. V. Corbett and C. A. Hurst. Are wave functions uniquely determined by their position and momentum distributions? *J. Austral. Math. Soc. Ser. B*, 20(2):182–201, 1977/78.

[CL86] Michael G. Crandall and Pierre-Louis Lions. On existence and uniqueness of solutions of hamilton-jacobi equations. *Nonlinear Analysis: Theory, Methods & Applications*, 10(4):353–370, 1986.

[CMS97] P. Cannarsa, A. Mennucci, and C. Sinestrari. Regularity results for solutions of a class of Hamilton-Jacobi equations. *Arch. Rational Mech. Anal.*, 140(3):197–223, 1997.

[Cor06] J. V. Corbett. The Pauli problem, state reconstruction and quantum-real numbers. *Rep. Math. Phys.*, 57(1):53–68, 2006.

[CP20a] P. Cardaliaguet and A. Porretta. An introduction to mean field game theory. In *Mean field games*, volume 2281 of *Lecture Notes in Math.*, pages 1–158. Springer, Cham, [2020] ©2020.

[CP20b] P. Cardaliaguet and A. Porretta. An introduction to mean field game theory. In *Mean field games*, volume 2281 of *Lecture Notes in Math.*, pages 1–158. Springer, Cham, [2020] ©2020.

[CS87] P. Cannarsa and H.M. Soner. On the singularities of the viscosity solutions to Hamilton-Jacobi-Bellman equations. *Indiana Univ. Math. J.*, 36(3):501–524, 1987.

[CS89] P. Cannarsa and H.M. Soner. Generalized one-sided estimates for solutions of Hamilton-Jacobi equations and applications. *Nonlinear Anal.*, 13(3):305–323, 1989.

[CS04] P. Cannarsa and C. Sinestrari. *Semiconcave functions, Hamilton-Jacobi equations, and optimal control*, volume 58 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser Boston, Inc., Boston, MA, 2004.

[CS22a] P. Cardaliaguet and P. Souganidis. Monotone solutions of the master equation for mean field games with idiosyncratic noise. *SIAM J. Math. Anal.*, 54(4):4198–4237, 2022.

[CS22b] P. Cardaliaguet and P.E. Souganidis. On first order mean field game systems with a common noise. *Ann. Appl. Probab.*, 32(3):2289–2326, 2022.

[CSS22] P. Cardaliaguet, B. Seeger, and P.E. Souganidis. Mean field games with common noise and degenerate idiosyncratic noise. *arXiv:2207.10209*, 2022.

[CY09] P. Cannarsa and Y. Yu. Singular dynamics for semiconcave functions. *J. Eur. Math. Soc. (JEMS)*, 11(5):999–1024, 2009.

[DLR19] F. Delarue, D. Lacker, and K. Ramanan. From the master equation to mean field game limit theory: a central limit theorem. *Electron. J. Probab.*, 24:Paper No. 51, 54, 2019.

[DLR20] F. Delarue, D. Lacker, and K. Ramanan. From the master equation to mean field game limit theory: large deviations and concentration of measure. *Ann. Probab.*, 48(1):211–263, 2020.

[Eva98] Lawrence C. Evans. *Partial differential equations*, volume 19. Providence, RI: American Mathematical Society, 1998.

[GM22a] W. Gangbo and A.R. Mészáros. Global well-posedness of master equations for deterministic displacement convex potential mean field games. *Comm. Pure Appl. Math.*, 75(12):2685–2801, 2022.

[GM22b] P.J. Graber and A.R. Mészáros. On mean field games and master equations through the lens of conservation laws. *arXiv:2208.10360*, 2022.

[GM22c] P.J. Graber and A.R. Mészáros. On mean field games and master equations through the lens of conservation laws. *arXiv:2208.10360*, 2022.

[GM23] P.J. Graber and A.R. Mészáros. On monotonicity conditions for mean field games. *J. Funct. Anal.*, 285(9):110095, 2023.

[GMMZ22] W. Gangbo, A.R. Mészáros, C. Mou, and J. Zhang. Mean field games master equations with nonseparable Hamiltonians and displacement monotonicity. *Ann. Probab.*, 50(6):2178–2217, 2022.

[Goe05a] R. Goebel. Convex optimal control problems with smooth Hamiltonians. *SIAM J. Control Optim.*, 43(5):1787–1811, 2005.

[Goe05b] R. Goebel. Duality and uniqueness of convex solutions to stationary Hamilton-Jacobi equations. *Trans. Amer. Math. Soc.*, 357(6):2187–2203, 2005.

[GR02] R. Goebel and R. T. Rockafellar. Generalized conjugacy in Hamiltonian-Jacobi theory for fully convex Lagrangians. volume 9, pages 463–473. 2002. Special issue on optimization (Montpellier, 2000).

[GS15] W. Gangbo and A. Święch. Existence of a solution to an equation arising from the theory of mean field games. *J. Differential Equations*, 259(11):6573–6643, 2015.

[GT19] W. Gangbo and A. Tudorascu. On differentiability in the Wasserstein space and well-posedness for Hamilton-Jacobi equations. *J. Math. Pures Appl. (9)*, 125:119–174, 2019.

[Hal13] Brian C. Hall. *Quantum Theory for Mathematicians*. Number 267 in Graduate Texts in Mathematics. Springer New York, 2013.

[HMC06] M. Huang, R.P. Malhamé, and P.E. Caines. Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the nash certainty equivalence principle. *Commun. Inf. Syst.*, 6(3):221–252, 2006.

[JR23] E.R. Jakobsen and A. Rutkowski. The master equation for mean field game systems with fractional and nonlocal diffusions. *arXiv:2305.18867*, 2023.

[Lio12a] P.-L. Lions. Théorie des jeux de champ moyen et applications. *Cours au Collège de France*, 2007-2012.

[Lio12b] P.-L. Lions. Cours au Collège de France, http://www.college-de-france. 2008–2012.

[LL07] J.-M. Lasry and P.-L. Lions. Mean field games. *Jpn. J. Math.*, 2(1):229–260, 2007.

[LLS22] J.-M. Lasry, P.-L. Lions, and B. Seeger. Dimension reduction techniques in deterministic mean field games. *Comm. Partial Differential Equations*, 47(4):701–723, 2022.

[May20] S. Mayorga. Short time solution to the master equation of a first order mean field game. *J. Differential Equations*, 268(10):6251–6318, 2020.

[McC97] R.J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128(1):153–179, 1997.

[MM24] A.R. Mészáros and C. Mou. Mean field games systems under displacement monotonicity. *SIAM J. Math. Anal.*, 56(1):529–553, 2024.

[MZ] C. Mou and J. Zhang. Wellposedness of second order master equations for mean field games with nonsmooth data. *Mem. Amer. Math. Soc., to appear*.

[MZ22a] C. Mou and J. Zhang. Mean field game master equations with anti-monotonicity conditions. *arXiv:2201.10762*, 2022.

[MZ22b] C. Mou and J. Zhang. Mean field games of controls: propagation of monotonicities. *Probab. Uncertain. Quant. Risk*, 7(3):247–274, 2022.

[MZ24] C. Mou and J. Zhang. Mean field game master equations with anti-monotonicity conditions. *J. Eur. Math. Soc. (JEMS), to appear*, arXiv:2201.10762, 2024+.

[Roc70a] R. T. Rockafellar. Conjugate convex functions in optimal control and the calculus of variations. *J. Math. Anal. Appl.*, 32:174–222, 1970.

[Roc70b] R.T. Rockafellar. *Convex analysis*. Princeton Mathematical Series. Princeton University Press, Princeton, N. J., 1970.

[Roc70c] R.T. Rockafellar. Generalized Hamiltonian equations for convex problems of Lagrange. *Pacific J. Math.*, 33:411–427, 1970.

[Vil03] Cédric Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[Vil09] Cédric Villani. *Optimal transport: Old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.

[Yu06] Y. Yu. A simple proof of the propagation of singularities for solutions of Hamilton-Jacobi equations. *Ann. Sc. Norm. Super. Pisa Cl. Sci. (5)*, 5(4):439–444, 2006.

[Zha17] Jianfeng Zhang. *Backward stochastic differential equations*, volume 86 of *Probability Theory and Stochastic Modelling*. Springer, New York, 2017. From linear to fully nonlinear theory.