MEAN FIELD GAMES WITH COMMON NOISE AND DEGENERATE IDIOSYNCRATIC NOISE

BY PIERRE CARDALIAGUET[1,a],
BENJAMIN SEEGER[2,b] AND PANAGIOTIS SOUGANIDIS[3,c]

[1]*CEREMADE, Université Paris–Dauphine,* [a]*cardaliaguet@ceremade.dauphine.fr*

[2]*Department of Statistics and Operations Research, University of North Carolina at Chapel Hill,* [b]*bseeger@unc.edu*

[3]*Department of Mathematics, University of Chicago,* [c]*souganidis@math.uchicago.edu*

We study the forward-backward system of stochastic partial differential equations describing a mean field game for a large population of small players subject to both idiosyncratic and common noise. The unique feature of the problem is that the idiosyncratic noise coefficient may be degenerate, so that the system does not admit smooth solutions in general. We develop a new notion of weak solutions for backward stochastic Hamilton-Jacobi-Bellman equations, and use this to build probabilistically weak solutions of the mean field game system. Under an additional monotonicity assumption, we prove the uniqueness of a strong solution.

1. Introduction. We investigate a notion of weak solution for the following stochastic Mean Field Games (MFG) system

$$(1.1) \quad \begin{cases} du_t = \big[-\beta\Delta u_t - \mathrm{tr}(a_t(x,m_t)D^2 u_t) + H_t(x, Du_t, m_t) - 2\beta\,\mathrm{div}(v_t)\big]dt \\ \qquad\qquad + v_t \cdot \sqrt{2\beta}dW_t \ \text{ in } \ [0,T)\times\mathbb{R}^d, \\ dm_t = \mathrm{div}\big[\beta Dm_t + \mathrm{div}(a_t(x,m_t)m_t) + m_t D_p H_t(x, Du_t, m_t)\big]dt \\ \qquad\qquad - \mathrm{div}(m_t\sqrt{2\beta}dW_t) \ \text{ in } \ (0,T]\times\mathbb{R}^d, \\ u_T(x) = G(x, m_T) \ \text{ and } \ m_0 = \bar{m}_0 \ \text{ in } \ \mathbb{R}^d, \end{cases}$$

which describes a differential game with infinitely many small players subject to a common noise, more precisely a Wiener process W with constant volatility $\beta > 0$, and an inhomogeneous idiosyncratic noise, the latter leading to the terms involving the possibly degenerate diffusion matrix a. The system consists of a backward stochastic Hamilton-Jacobi-Bellman (HJB) equation satisfied by the value function u of a typical player, and a forward Fokker-Plank (FP) describing the evolution of the population density m. The third unknown, the $\mathbb{R}^d$-valued map v in the backward HJB equation, ensures that u is adapted to the filtration generated by the common noise W. The Hamiltonian $H : [0,T]\times\mathbb{R}^d\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$ (where $\mathcal{P}(\mathbb{R}^d)$ is the set of Borel probability measures on $\mathbb{R}^d$) is continuous and convex in the gradient variable, the map $a : [0,T]\times\mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d) \to \mathbb{R}^{d\times d}$ is continuous with values in the set of symmetric nonnegative matrices, and $G : \mathbb{R}^d\times\mathcal{P}(\mathbb{R}^d)$ is the terminal cost. More precise assumptions are listed later.

In [8], Cardaliaguet, Delarue, Lasry and Lions proved the existence and uniqueness of classical solutions of (1.1) set on the torus for a nondegenerate matrix a, H with a separated dependence on Du and m, and strictly monotone coupling functions. Cardaliaguet and

Souganidis considered in [10] the first-order version of (1.1) set in the whole space when there is no idiosyncratic noise, that is, $a \equiv 0$ and $\beta > 0$.

The goal of the present note is to extend the results of [8] and [10] to the whole space with nonzero but possibly degenerate a. As in [10], the main feature of the problem that distinguishes it from the existing literature on MFG systems with common noise is the nonexistence of classical solutions in general, a fact that necessitates the introduction of a suitable notion of weak solution. Due to the second-order nature of the system, the methods of [10] are no longer applicable, and must be substantially revisited.

The starting point in the analysis is the observation that, because all players are affected by the same noise W, the stochasticity can be cancelled out by shifting the density m of the players along this noise. Indeed performing the change of variables

$$(1.2) \qquad \widetilde{u}_t(x) = u_t(x + \sqrt{2\beta}W_t) \quad and \quad \widetilde{m}_t = (\mathrm{Id} - \sqrt{2\beta}W_t)_\sharp m_t$$

and setting, for $(t, x, p, m) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$,

$$(1.3) \qquad \widetilde{a}_t(x, m) = a_t(x + \sqrt{2\beta}W_t, (\mathrm{Id} + \sqrt{2\beta}W_t)_\sharp m),$$

$$(1.4) \qquad \widetilde{H}_t(x, p, m) = H_t(x + \sqrt{2\beta}W_t, p, (\mathrm{Id} + \sqrt{2\beta}W_t)_\sharp m),$$

$$(1.5) \qquad \widetilde{G}(x, m) = G(x + \sqrt{2\beta}W_T, (\mathrm{Id} + \sqrt{2\beta}W_T)_\sharp m),$$

and

$$(1.6) \qquad \widetilde{v}_t(x) = v_t(x - \sqrt{2\beta}W_t) - \sqrt{2\beta}Du_t(x - \sqrt{2\beta}W_t), \quad and \quad M_t = \int_0^t \widetilde{v}_s dW_s,$$

leads, at least formally, to the system

$$(1.7) \quad \begin{cases} d\widetilde{u}_t = \left[-\operatorname{tr}(\widetilde{a}_t(x, \widetilde{m}_t)D^2\widetilde{u}_t) + \widetilde{H}_t(x, D\widetilde{u}_t, \widetilde{m}_t) \right] dt + dM_t & \text{in } [0, T) \times \mathbb{R}^d, \\ d\widetilde{m}_t = \operatorname{div}\left[\operatorname{div}(\widetilde{a}_t(x, \widetilde{m}_t)\widetilde{m}_t) + \widetilde{m}_t D_p\widetilde{H}_t(x, D\widetilde{u}_t, \widetilde{m}_t) \right] dt & \text{in } (0, T] \times \mathbb{R}^d, \\ \widetilde{u}_T(x) = \widetilde{G}(x, \widetilde{m}_T), \quad \widetilde{m}_0 = \bar{m}_0 \text{ in } \mathbb{R}^d. \end{cases}$$

Observe that the new functions $\widetilde{a}$, $\widetilde{H}$, and $\widetilde{G}$ are random and adapted to the common noise W. Note also that, since they arise from a, H, and G through translations along W, they retain whatever regularity and structural conditions were satisfied by the deterministic data uniformly over $m \in \mathcal{P}$ (see Remark 1.1 below).

Crucially, the new forward FP equation in (1.7) is no longer stochastic, and, therefore, can be interpreted in the usual sense of distributions. On the other hand, the backward equation for $\widetilde{u}$ still involves a stochastic correction term, namely, the martingale M.

1.1. *Informal discussion of the main results.*

The first main task of the paper is to develop a useful notion of weak solution for backward stochastic HJB equations of the form

$$(1.8) \quad \begin{cases} d\widetilde{u}_t = \left[-\operatorname{tr}(\widetilde{a}_t(x)D^2\widetilde{u}_t) + \widetilde{H}_t(x, D\widetilde{u}_t) \right] dt + dM_t & \text{in } [0, T) \times \mathbb{R}^d \quad and \\ \widetilde{u}_T = \widetilde{G} \text{ in } \mathbb{R}^d, \end{cases}$$

where the unknowns $(\widetilde{u}, M)$ and the data $\widetilde{a}$, $\widetilde{H}$, and $\widetilde{G}$ are adapted to a common filtration and M is a martingale taking values in a space of generalized functions (distributions).

Equations like (1.8) should typically be satisfied by the value function of optimal control problems with random coefficients adapted to a given filtration, as studied by Peng [31], who

proved existence and uniqueness of solutions for smooth data and uniformly elliptic $\widetilde{a}$. They can also be studied with the theory of path-dependent viscosity solutions, a notion which involves taking derivatives on the path space; see for instance Qiu [32] and Qiu and Wei [33]. This theory allows for the treatment of more general equations than (1.8), for example, with non-constant volatility β, and, other than standard growth and regularity assumptions, few conditions are required for the Hamiltonian. However, in our context, in order to study the full system (1.7), in particular, the forward FP equation, an understanding of the gradient $D\widetilde{u}_t$ is required, which is not accessible in the path-dependent theory.

The strategy of [10] for the case $\widetilde{a} \equiv 0$ is to adapt the classical approach of Kružhkov [23] for forward Hamilton-Jacobi equations (see also Douglis [16], Evans [17, Section 3.3], and Fleming [20]), in which solutions are understood in the almost-everywhere sense, and well-posedness is established in the class of functions that are globally semiconcave in the state variable. In the setting of [10], this means that the equation is understood as a stochastic evolution satisfied for a.e. $x \in \mathbb{R}^d$, with the martingale M taking values in $L^\infty(\mathbb{R}^d)$.

There are immediate obstacles in adapting these methods to the second-order equation (1.8). As noted above, one cannot expect $\widetilde{u}$ to possess any additional regularity than in the first-order setting. In fact, the inherently first-order arguments employed in [23, 16, 17, 20, 10] do not carry over to (1.8), whose solutions no longer have a finite domain of dependence. In addition, (1.8) can no longer be interpreted in an almost-everywhere sense, because the distribution $\mathrm{tr}(\widetilde{a}_t D^2 \widetilde{u}_t)$ is not a function in general.

Instead, here we interpret (1.8) as a stochastic evolution in the space of distributions, or, more precisely, locally finite measures (see Definition 3.1 below). Of central importance to the well-posedness, just as in the first-order setting, is a global semiconcavity bound for the solutions. The existence of such a solution (Theorem 3.1) relies on the propagation in time of semiconcavity for deterministic HJB equations (that is, (1.8) with $M \equiv 0$). This propagation estimate (Proposition 2.1 below), as far as we know, does not appear in the literature in the present generality.

The uniqueness and stability of solutions (Theorems 3.2 and 3.3) rely on the existence of a smooth, positive supersolution ψ of the forward-in-time, deterministic linearization of (1.8), namely

$$\partial_t \psi - \mathrm{tr}[A\psi] + B \cdot D\psi \geq 0 \quad \text{in } \mathbb{R}^d \times [0, T],$$

which holds uniformly over bounded $A \in \mathbb{S}^d$ and $B \in \mathbb{R}^d$. When $A \equiv 0$, it is enough to take the indicator function of a ball with arbitrarily large radius shrinking at a fixed rate. For $A \neq 0$, it is possible to find such a ψ with exponential decay at infinity, but care must be taken when using it as a test function, due to the fact that $D^2 u$ is only locally bounded as a measure.

Once the HJB equation (1.8) is well understood, we turn to the study of the system (1.7). The existence and uniqueness of a probabilistically strong solution, that is, adapted to the filtration of the common noise W, was proved directly in [10] in the first-order case $\widetilde{a} \equiv 0$ under the additional assumption that the Hamiltonian is of "separated" form, that is,

$$H_t(x, p, m) = \hat{H}_t(x, p) - F_t(x, m),$$

with the coupling functions F and G strictly monotone. This latter assumption allows for a direct proof of the uniqueness of the solution using the well-known "duality trick" of Lasry and Lions [26].

Our second main contribution is to generalize the existence result by not only allowing for a degenerate diffusion matrix, but also imposing only mild assumptions on the joint dependence of a, H, G on u and m. In particular, under mild regularity assumptions with respect to the measure variable m, we prove in Theorem 4.1 the existence of global, probabilistically

weak solutions of (1.7). This is achieved by interpreting (1.7) as a fixed point problem for $\widetilde{m}$ in a subset of $L^\infty(\Omega, C([0,T], \mathcal{P}_2(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)))$, where $\mathcal{P}_2$ is the space of probability measures with finite second moment. The main issue with this strategy is the failure of compactness, due to the fact that the probability space Ω may be quite large. Compactness is restored by studying a variant of (1.7) in which the probability space is discretized in an appropriate way. After invoking Kakutani's fixed point theorem, a weak solution is obtained upon taking weak limits.

The proof of Theorem 4.1 borrows several ideas from the work of Carmona, Delarue, and Lacker [12], who study mean field games with common noise from a purely probabilistic perspective. The main difference with our work is that we proceed at the PDE level and not in the optimal control formulation, as we exploit the regularity properties of the value function in the state variable x. We therefore bypass concerns about the existence or uniqueness of optimal trajectories, instead developing an understanding of the PDE system itself.

The final main result of the paper (Theorem 4.2) is the uniqueness of solutions of (1.7) under a monotonicity assumption. More precisely, borrowing terminology from the theory of stochastic differential equations, it is shown that pathwise uniqueness holds, and thus, by a Yamada-Watanabe-type argument, the unique solution is in fact strong in the probabilistic sense, which means not only that $(\widetilde{u}, M, \widetilde{m})$ is adapted with respect to the filtration generated by the common noise W, but also, can be represented as a measurable functional of the Wiener process W.

The uniqueness proof, whose essential method was introduced in [26], involves studying the time evolution of the quantity

$$t \mapsto \int_{\mathbb{R}^d} (\widetilde{u}_t^1(x) - \widetilde{u}_t^2(x))(\widetilde{m}_t^1(x) - \widetilde{m}_t^2(x))dx$$

for two solutions $(\widetilde{u}^1, \widetilde{m}^1)$ and $(\widetilde{u}^2, \widetilde{m}^2)$. For the particular system (1.7), such an argument cannot proceed as stated, because neither $\widetilde{m}$ nor $\widetilde{u}$ is regular enough to be used as a test function in each others' equations. The resulting analysis, which involves estimating commutators after mollifying and localizing, relies precisely on the fact that $D^2\widetilde{u}$ is a locally finite measure for a solution $\widetilde{u}$ of the backward stochastic HJB equation.

1.2. *Previous work on MFG.* The MFG system, in the absence of common noise, was introduced and studied by Lasry and Lions in [26, 24, 25]. In the presence of both idiosyncratic and common noise, the stochastic MFG system was first investigated in [8] in which the state space is the torus $\mathbb{R}^d/\mathbb{Z}^d$, and the existence and uniqueness of a strong solution was established for separated Hamiltonians, strictly monotone coupling functions, and nondegenerate diffusion. The result was extended to $\mathbb{R}^d$ by Carmona and Delarue in the monograph [11].

An alternative analytic approach to study MFG equilibria with a common noise is the master equation, introduced by Lasry and Lions and presented by Lions in [27], which is a deterministic, nonlinear, nonlocal transport equation in the space of probability measures. The existence and uniqueness of classical solutions to the master equation was first established in [8] (see also [11]). A related approach, presented in [27], is to lift probability measures to the Hilbert space of L^2-integrable random variables via their probability laws.

The master equation has recently attracted much attention, especially for problems with common noise. In [7], Cardaliaguet, Cirant and Porretta build solutions using a splitting method. The setting of a finite state space is investigated by Bertucci, Lasry and Lions [4, 5] and Bayraktar, Cecchin, Cohen and Delarue [1]; in the latter work, it is also demonstrated that common noise can restore uniqueness in the absence of monotonicity. Another approach to uniqueness is explored by Gangbo, Mészáros, Mou and Zhang [21] via displacement convexity. Various notions of weak solutions to master equations are proposed by Mou and

Zhang [30] and Bertucci [2, 3]. Through a combination of the Hilbertian interpretation and Bertucci's notion of monotone solutions, Cardaliaguet and Souganidis build a unique, global, weak solution to the master equation with common noise but without idiosyncratic noise [9]; we note that this approach does not seem to adapt to problems with idiosyncratic noise.

The analysis of MFG equilibria with a common noise can also be viewed from a purely probabilistic viewpoint. This idea was first developed in [12], and indeed, our analysis of probabilistically weak solutions of the MFG system borrows several ideas from this paper. The existence results of [12] were recently extended by Djete [15] to MFG problems with an interaction through the controls.

1.3. *Main assumptions.* Throughout the paper, we introduce diffusion coefficients, Hamiltonians, and terminal data that, depending on the setting, take on different forms (random or deterministic, coupling within the forward-backward system, monotonicity, etc.). We collect and discuss here the basic assumptions that are used in all sections.

The degenerate diffusion matrices will always take the form

$$(1.9) \qquad \mathcal{A} := \Sigma\Sigma^\tau, \ \ \text{where, for some } m \le d, \ \Sigma \in C([0,T]; C^{1,1}(\mathbb{R}^d; \mathbb{R}^{d\times m})).$$

The terminal data will be a given function $\mathcal{G}$ such that

$$(1.10) \qquad \mathcal{G} : \mathbb{R}^d \to \mathbb{R} \ \ \text{is globally bounded, Lipschitz, and semiconcave.}$$

Finally, the Hamiltonian $\mathcal{H}$ is assumed to have the following regularity and structural properties:

$$(1.11) \quad \begin{cases} \text{for all } t \in [0,T], \ \mathcal{H}_t \in C^2(\mathbb{R}^d \times \mathbb{R}^d); \\[1mm] \text{for all } R > 0, \ \mathbb{R}^d \times B_R \ni (x,p) \mapsto \mathcal{H}_t(x,p) \\[1mm] \qquad \text{is uniformly bounded and Lipschitz for } t \in [0,T]; \\[1mm] [0,T] \ni t \mapsto \mathcal{H}_t(x,p) \ \ \text{is continuous uniformly over } (x,p) \in \mathbb{R}^d \times B_R; \text{ and} \\[1mm] p \mapsto \mathcal{H}_t(x,p) \ \ \text{is uniformly convex,} \end{cases}$$

and

$$(1.12) \quad \begin{cases} \text{for some } \lambda_0, C_0 > 0 \text{ and all } t \in [0,T], \ p,q \in \mathbb{R}^d, \text{ and } |z| = 1, \\[1mm] |D_x \mathcal{H}_t(x,p)| \le C_0 + \lambda_0(p \cdot D_p \mathcal{H}_t(x,p) - \mathcal{H}_t(x,p)) \text{ and} \\[1mm] D_{pp}^2 \mathcal{H}_t(x,p) q \cdot q + 2D_{px}^2 \mathcal{H}_t(x,p) z \cdot q + D_{xx}^2 \mathcal{H}_t(x,p) z \cdot z \\[1mm] \qquad + \lambda_0 \left(D_p \mathcal{H}_t(x,p) \cdot p - \mathcal{H}_t(x,p)\right) \ge -C_0. \end{cases}$$

Condition (1.12), which appears also in [10, (1.10)(iii)], is an assumption on the nature of the coupling between the x and p variables in $\mathcal{H}$. It also has a natural interpretation from the viewpoint of the underlying stochastic control problem. Indeed, (1.12) is equivalent to the following condition on the Legendre transform $\mathcal{H}_t^*(x,\alpha) = \sup_p(p \cdot \alpha - \mathcal{H}_t(x,p))$:

$$(1.13) \quad \begin{cases} \text{for all } t \in [0,T] \text{ and } x, \alpha \in \mathbb{R}^d, \\[1mm] |D_x \mathcal{H}_t^*(x,\alpha)| \le C_0 + \lambda_0 \mathcal{H}_t^*(x,\alpha) \ \ \text{and} \\[1mm] D_{xx}^2 \mathcal{H}_t^*(x,\alpha) \le (C_0 + \lambda_0 \mathcal{H}_t^*(x,\alpha)) \operatorname{Id}. \end{cases}$$

The formula for the value function solving the terminal value problem

$$-\partial_t U - \operatorname{tr}[\mathcal{A}_t(x) D^2 U_t(x)] + \mathcal{H}_t(x, DU_t(x)) = 0 \ \ \text{in } (0,T) \times \mathbb{R}^d, \quad U_T = \mathcal{G}$$

is

(1.14)
$$(t,x) \mapsto \inf_\alpha \mathbb{E}\left[\int_t^T \mathcal{H}_s^*(X_s^{\alpha,t,x}, \alpha_s)ds + \mathcal{G}(X_T^{\alpha,t,x})\right],$$

where, for some m-dimensional Brownian motion B, $X^{\alpha,t,x}$ solves

$$\begin{cases} dX_s^{\alpha,t,x} = \alpha_s ds + \Sigma_s(X_s^{\alpha,t,x})dB_s, & s \in [t,T], \\ X_t^{\alpha,t,x} = x. \end{cases}$$

The infimum in (1.14) is taken over progressively measurable controls $\alpha : [t,T] \to \mathbb{R}^d$ that satisfy the admissibility criterion

$$\mathbb{E}\left[\int_t^T \sup_{x \in \mathbb{R}^d} \mathcal{H}_s^*(x, \alpha_s)ds\right] < \infty.$$

Formally differentiating (1.14) twice in the x-variable, one arrives at a semiconcavity bound precisely when (1.9) and (1.10) hold, and when

$$\mathbb{E}\left[\int_t^T \sup_{x \in \mathbb{R}^d} |D_x \mathcal{H}_s^*(x, \alpha_s)|ds + \int_t^T \sup_{x \in \mathbb{R}^d, |z|=1} D_{xx}^2 \mathcal{H}_s^*(x, \alpha_s)z \cdot zds\right] < \infty.$$

Then (1.13) is exactly what one needs in order for the latter to hold.

This demonstrates that (1.12) is the most natural assumption to make on $\mathcal{H}$ in order for semiconcavity to hold for the value function U. These arguments can all be made rigorous, but we choose instead to prove Proposition 2.1 below by performing the differentiation on the level of the PDE, keeping with the theme of this paper.

REMARK 1.1. Observe that, if the functions H, G, and a in the original system (1.1) satisfy assumptions (1.9) - (1.12) uniformly over $m \in \mathcal{P}(\mathbb{R}^d)$, then so do the transformed functions $\widetilde{H}$, $\widetilde{G}$, and $\widetilde{a}$ defined via (1.3), (1.4), and (1.5), uniformly over the probability space on which the Wiener process W is defined. This is a feature of the fact that the volatility β is constant in the state variable. This is not merely a simplification: the precise propagation of semiconcavity proved in Proposition 2.1 below depends on the various constants in (1.9) - (1.12) above.

1.4. *Organization of the paper.* In Section 2, we prove the propagation of semiconcavity estimate for backward deterministic HJB equations, and also list some properties of semiconcave functions that are used throughout the paper. Section 3 introduces the definition of solutions for the backward stochastic HJB equation (1.8), including a discussion of stochastic processes taking values in the space of locally finite measures, and establishes well-posedness results. Section 4 involves the study of weak and strong solutions to the MFG system (1.7).

1.5. *Notation.* Given an open or closed domain U in Euclidean space, $\mathcal{B}_U$ denotes the Borel σ-algebra inside U and $C(U)$ is the space of continuous functions on U endowed with the topology of (local) uniform convergence, and $C_c(U)$ is the space of compactly supported continuous functions on U. For a function f defined on a measure space $\mathcal{A}$, we set $\|f\|_{\infty,\mathcal{A}} :=$ ess $\sup_{x \in \mathcal{A}} |f(x)|$, and, where it does not create confusion, we simply write $\|f\|_\infty = \|f\|_{\infty,\mathcal{A}}$. Throughout the paper, (generalized) function spaces $X(\mathbb{R}^d)$ on all of $\mathbb{R}^d$ are simply denoted by X when this does not cause confusion.

We define $\mathcal{D} = C_c^\infty$, and the space $\mathcal{D}'$ of distributions consists of real-valued linear functionals $\mathcal{T} : \mathcal{D} \to \mathbb{R}$ satisfying the following local continuity property:

$$\text{if } \{(\phi_n)_{n \in \mathbb{N}}, \phi\} \subset C_c^\infty, \quad \bigcup_{n \in \mathbb{N}} \operatorname{supp} \phi_n \text{ is bounded, and}$$

$$D^k \phi_n \xrightarrow{n \to \infty} D^k \phi \text{ uniformly for all } k = 0, 1, 2, \ldots, \text{ then}$$

$$\mathcal{T}(\phi_n) \xrightarrow{n \to \infty} \mathcal{T}(\phi),$$

where $D^k \phi$ refers to the tensor of all partial derivatives of ϕ of order k. We often write $\mathcal{T}(\phi) = \langle \mathcal{T}, \phi \rangle$.

The space of locally finite, signed, Radon measures on $\mathbb{R}^d$ is denoted by $\mathcal{M}_{\mathrm{loc}} = \mathcal{M}_{\mathrm{loc}}(\mathbb{R}^d)$. For $\mu \in \mathcal{M}_{\mathrm{loc}}$ and $f \in C_c$, $\langle \mu, f \rangle := \int_{\mathbb{R}^d} f(x) \mu(dx)$, and, for an open bounded set $U \subset \mathbb{R}^d$, we write the total variation norm as

$$\|\mu\|_{TV(U)} := \sup_{f \in C_c(U), \|f\|_\infty \leq 1} \langle \mu, f \rangle.$$

We define the quasi-norm

$$\|\mu\|_{\mathcal{M}_{\mathrm{loc}}} := \sum_{n=1}^\infty \|\mu\|_{TV(B_n)} \wedge 2^{-n},$$

where B_R is the open ball centered at 0 and of radius R. We will say $\mathcal{A} \subset \mathcal{M}_{\mathrm{loc}}$ is bounded if $\{\|\mu\|_{\mathcal{M}_{\mathrm{loc}}}\}_{\mu \in \mathcal{A}}$ is bounded, and, for $p \in [1, \infty]$ and a measure space $(\mathcal{A}, \gamma)$,

$$L^p(\mathcal{A}, \mathcal{M}_{\mathrm{loc}}) := \left\{ \mu : \mathcal{A} \to \mathcal{M}_{\mathrm{loc}} \mid \|\mu\|_{L^p(\mathcal{A}, \mathcal{M}_{\mathrm{loc}})} := \left(\int_{\mathcal{A}} \|\mu(a)\|_{\mathcal{M}_{\mathrm{loc}}}^p \gamma(da) \right)^{1/p} < \infty \right\},$$

with the usual modification if $p = \infty$. The metric $(\mu, \nu) \mapsto \|\mu - \nu\|_{\mathcal{M}_{\mathrm{loc}}}$ makes $\mathcal{M}_{\mathrm{loc}}$ into a complete metric space with the topology of local convergence in total variation, and, similarly, $L^p(\mathcal{A}, \mathcal{M}_{\mathrm{loc}})$ is a metric space with the metric $(\mu, \nu) \mapsto \|\mu - \nu\|_{L^p(\mathcal{A}, \mathcal{M}_{\mathrm{loc}})}$.

The Sobolev space of functions $f \in L^1$ for which the distributional derivative Df belongs to L^1 is denoted $W^{1,1}$. Its dual is a space of distributions:

$$W^{-1,\infty} := \left\{ \mathcal{T} \in \mathcal{D}' : \|\mathcal{T}\|_{W^{-1,\infty}} := \sup_{\phi \in \mathcal{D}} \frac{|\mathcal{T}(\phi)|}{\|\phi\|_{W^{1,1}}} < \infty \right\}.$$

If $f \in L^\infty$, then the distributional derivative Df belongs to $W^{-1,\infty}(\mathbb{R}^d)$, and, more generally, $W^{-1,\infty}(\mathbb{R}^d)$ consists of distributions of the form $f_0 + Df_1$ for $f_0, f_1 \in L^\infty$, and, in fact,

$$\|\mathcal{T}\|_{W^{-1,\infty}} := \inf \left\{ \|f_0\|_\infty + \|f_1\|_\infty : \mathcal{T} = f_0 + Df_1, \, f_0, f_1 \in L^\infty \right\}.$$

The space of probability measures on $\mathbb{R}^d$ is denoted by $\mathcal{P}(\mathbb{R}^d)$, and, for $p > 0$,

$$\mathcal{P}_p(\mathbb{R}^d) := \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) : \int_{\mathbb{R}^d} |x|^p \mu(dx) < \infty \right\}.$$

The p-Wasserstein distance $\mathbf{d}_p$ on $\mathcal{P}_p(\mathbb{R}^d)$ is defined by

$$\mathbf{d}_p(\mu, \mu') := \inf_\gamma \left(\iint_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^p d\gamma(x, y) \right)^{1/p},$$

where the infimum is taken over probability measures $\gamma \in \mathcal{P}_p(\mathbb{R}^d \times \mathbb{R}^d)$ with first and second marginal respectively μ and μ'. Given a probability space $(\Omega, \mathbb{P})$, we write $\mathbb{E}_\mathbb{P} := \int_\Omega d\mathbb{P}$.

Throughout, we fix $\rho \in C^2(\mathbb{R}^d)$ satisfying $\rho \geq 0$, $\operatorname{supp} \rho \subset B_1$, and $\int_{\mathbb{R}^d} \rho = 1$, and, for $\delta > 0$ and $x \in \mathbb{R}^d$, we set

$$(1.15) \qquad \rho_\delta(x) := \frac{1}{\delta^d} \rho\left(\frac{x}{\delta}\right).$$

The transpose of a matrix M is denoted by M^τ. We define, for $X \in \mathbb{S}^d$, the space of $d \times d$ symmetric matrices, the degenerate elliptic operator

$$m_+(X) := \max_{|v| \leq 1} Xv \cdot v.$$

Note that $m_+(X)$ is the maximum of 0 and the largest eigenvalue of X, and therefore m_+ is increasing with respect to the partial order on $\mathbb{S}^d$. The Lebesgue measure on $\mathbb{R}^d$ is denoted by $\mathcal{L}$.

Throughout proofs, we denote by C a positive constant that may change from line to line, depending only on the relevant data in the assumptions of the corresponding theorem/proposition/lemma.

2. Propagation of semiconcavity for deterministic HJB equations.

The purpose of this section is to prove that semiconcavity bounds propagate in time for solutions of the deterministic HJB equation

$$(2.1) \qquad \begin{cases} -\partial_t u_t - \operatorname{tr}[a_t(x) D^2 u_t] + H_t(x, Du_t(x)) = 0 & \text{in } [0, T] \times \mathbb{R}^d \quad \text{and} \\ u_T = G & \text{in } \mathbb{R}^d, \end{cases}$$

where a, H, and G are fixed and deterministic, and satisfy the assumptions from subsection 1.3.

Recall that a function $u : \mathbb{R}^d \to \mathbb{R}$ is said to be semiconcave with constant $C > 0$ if

$$u(x + h) + u(x - h) - 2u(x) \leq C|h|^2 \quad \text{for all } x, h \in \mathbb{R}^d.$$

The result below gives a precise estimate of the positive part of the maximal eigenvalue of the density of the measure $D^2 u$, which, in turn, controls the semiconcavity constant (see Lemma 2.1 and Remark 2.1 below).

PROPOSITION 2.1. *Assume a, G, and H satisfy* (1.9), (1.10), (1.11), *and* (1.12), *and let u be the unique viscosity solution of* (2.1). *Then, for all $t \in [0, T]$,*

$$(2.2) \qquad \|u_t\|_\infty \leq \|G\|_\infty + \|H.(\cdot, 0)\|_\infty (T - t).$$

Moreover, u is semiconcave and Lipschitz in the x-variable, uniformly on $[0, T] \times \mathbb{R}^d$, and there exists a constant $C > 0$ depending only on T and the various data in those assumptions such that, for all $t \in [0, T]$,

$$\operatorname*{ess\,sup}_{x \in \mathbb{R}^d} \left\{ \left(m_+\left(\frac{dD^2 u_t(x)}{d\mathcal{L}}\right)^2 + |Du_t(x)|^2 \right)^{1/2} - \sqrt{2}\lambda_0 u_t(x) \right\}_+$$

$$(2.3) \qquad \leq C(T - t)$$

$$+ e^{C(T-t)} \operatorname*{ess\,sup}_{x \in \mathbb{R}^d} \left\{ m_+\left(\frac{dD^2 g(x)}{d\mathcal{L}}\right)^2 + |Dg(x)|^2 \right)^{1/2} - \sqrt{2}\lambda_0 g(x) \right\}_+.$$

Finally, $-\operatorname{tr}[aD^2 u] \in L^\infty([0, T], \mathcal{M}_{\mathrm{loc}}(\mathbb{R}^d))$, and (2.1) *is satisfied in the distributional sense.*

The fact that solutions of HJB equations are semiconcave is well known, and is a consequence of the minimization problem that defines the value function u; see for instance [6, 22, 28]. Note that these references do not provide bounds as in Proposition 2.1 in the generality assumed here, since either the derivatives of H are assumed uniformly bounded, or an a priori bound on $\|Du\|_\infty$ is assumed. The main feature of Proposition 2.1 is the propagation of a bound for the quantity on the left-hand side of (2.3), which ultimately leads to the semiconcavity and Lipschitz bounds for solutions of the stochastic HJB equation (1.8).

We summarize some properties and results concerning semiconcave functions used in the proof of Proposition 2.1 and throughout the paper. For more details and explanations, see for instance the book of Evans and Gariepy [18].

LEMMA 2.1.   *Assume* $u : \mathbb{R}^d \to \mathbb{R}$ *is bounded and semiconcave, with semiconcavity constant* $C \geq 0$. *Then the following hold:*

(a) *The distribution* D^2u *is a measure of the form* $D^2u = f(x)d\mathcal{L} - d\mu$, *where* μ *is nonnegative and singular with respect to the Lebesgue measure* $\mathcal{L}$ *and* $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d; \mathbb{S}^d)$ *satisfies* $f \leq C\,\mathrm{Id}$ *Lebesgue-a.e.*
(b) *The function* u *is Lipschitz, and* $\|Du\|_\infty \leq 2\sqrt{C\,\|u\|_\infty}$. *In particular,* $D^2u \in W^{-1,\infty}$.
(c) *There exists a dimensional constant* $c_d > 0$ *such that* $\left\|D^2u\right\|_{\mathcal{M}_{\mathrm{loc}}(\mathbb{R}^d)} \leq c_d(\|Du\|_\infty + C)$.
(d) *If* $(u^n)_{n \in \mathbb{N}} : \mathbb{R}^d \to \mathbb{R}$ *is a sequence of globally Lipschitz, semiconcave functions, uniformly in* $n \in \mathbb{N}$, *and, as* $n \to \infty$, u^n *converges locally uniformly to* u, *then, as* $n \to \infty$, Du^n *converges almost everywhere to* Du.

REMARK 2.1.   The statement in part (a) is actually equivalent to semiconcavity. The semiconcavity constant for a globally semiconcave function u is equal to

$$\operatorname*{ess\,sup}_{x \in \mathbb{R}^d} m_+ \left(\frac{dD^2u}{d\mathcal{L}}(x) \right),$$

and, hence, the bound in Proposition 2.1 indeed propagates the semiconcavity estimate for u. Throughout the rest of the paper, where it does not create confusion, we abuse notation and denote by $\operatorname{ess\,sup} m_+(D^2u)$ the semiconcavity constant for a function $u : \mathbb{R}^d \to \mathbb{R}$.

Proposition 2.1, and in particular the estimate (2.3), is proved by differentiating the equation (2.1), and, as such, it is necessary to perform a regularization to make the arguments rigorous. For $0 < \varepsilon < 1$, set $a^\varepsilon = \sigma\sigma^\tau + \varepsilon\,\mathrm{Id}$ and $G^\varepsilon = G * \rho_\varepsilon$ for ρ_ε as in (1.15), and consider the classical solution u^ε of

(2.4)
$$\begin{cases} -\partial_t u^\varepsilon_t - \mathrm{tr}[a^\varepsilon_t(x)D^2u^\varepsilon_t(x)] + H_t(x, Du^\varepsilon_t(x)) = 0 & \text{in } [0,T] \times \mathbb{R}^d \quad \text{and} \\ u^\varepsilon_T = G^\varepsilon & \text{in } \mathbb{R}^d. \end{cases}$$

LEMMA 2.2.   *The solution* u^ε *of* (2.4) *satisfies* (2.2) *and* (2.3), *where the constant* $C > 0$ *is as in Proposition 2.1 and independent of* $\varepsilon > 0$.

We first prove Proposition 2.1 with Lemma 2.2 in hand.

PROOF OF PROPOSITION 2.1.   In view of Lemma 2.2, $(u^\varepsilon)_{0 < \varepsilon < 1}$ is uniformly bounded, Lipschitz, and semiconcave in the space variable. The stability properties of viscosity solutions imply that, as $\varepsilon \to 0$, u^ε converges locally uniformly in $[0,T] \times \mathbb{R}^d$ to u, and this then yields the Lipschitz regularity and semiconcavity of u.

As $\varepsilon \to 0$, D^2u^ε converges to D^2u in the sense of distributions, and, in view of Lemma 2.1(d), as $\varepsilon \to 0$, Du^ε converges almost everywhere to Du. We conclude that, as $\varepsilon \to 0$, the

left-hand side of (2.4) converges to the left-hand side of (2.1) in the sense of distributions. The claim about $-\operatorname{tr}[a_t(x)D^2u_t(x)]$ follows from Lemma 2.1(c).

Using Jensen's inequality we see that, for all $x \in \mathbb{R}^d$,

$$
\left(m_+ \left(D^2 G^\varepsilon(x) \right)^2 + |DG^\varepsilon(x)|^2 \right)^{1/2} - \sqrt{2}\lambda_0 G^\varepsilon(x)
$$

$$
\leq \rho_\varepsilon * \left\{ \left(m_+ \left(\frac{dD^2 G}{d\mathcal{L}} \right)^2 + |DG|^2 \right)^{1/2} - \sqrt{2}\lambda_0 G \right\}(x)
$$

$$
\leq \operatorname*{ess\,sup}_{x \in \mathbb{R}^d} \left\{ m_+ \left(\frac{dD^2 g(x)}{d\mathcal{L}} \right)^2 + |Dg(x)|^2 \right)^{1/2} - \sqrt{2}\lambda_0 g(x) \right\}_+ =: M.
$$

Rearranging terms in the bound (2.3) satisfied by u^ε gives, for all $t \in [0,T]$,

$$
D^2 u_t^\varepsilon \leq \left(\left[(C(T-t) + e^{C(T-t)}M + \sqrt{2}u_t^\varepsilon)^2 - |Du_t^\varepsilon|^2 \right]_+ \right)^{1/2} \operatorname{Id}.
$$

As $\varepsilon \to 0$, the left- and right-hand sides above converge in the sense of distributions, yielding, for all $t \in [0,T]$,

$$
D^2 u_t \leq \left(\left[(C(T-t) + e^{C(T-t)}M + \sqrt{2}u_t)^2 - |Du_t|^2 \right]_+ \right)^{1/2} \operatorname{Id}.
$$

In particular, the same bound holds for the density $\frac{dD^2 u}{d\mathcal{L}}$ a.e. in view of Lemma 2.1(a). Rearranging terms yields the desired bound (2.3). □

We now present the

PROOF OF LEMMA 2.2. Throughout the proof, we suppress the dependence on ε, but note that the solution u is smooth.

For $\xi, \eta \in \mathbb{R}^d$ and $f : \mathbb{R}^d \to \mathbb{R}$, we use the shorthand $f_\xi = \xi \cdot Df$ and $f_{\eta\eta} = D^2 f\eta \cdot \eta$. We also suppress dependence on (t,x) below where it does not create confusion, and note that the arguments of H and its derivatives will always be (x, Du).

Differentiating (2.4) twice leads to

(2.5) $\qquad -\partial_t u_\xi - \operatorname{tr}[aD^2 u_\xi] + D_p H \cdot Du_\xi - \operatorname{tr}[a_\xi D^2 u] + \xi \cdot D_x H_t = 0$

and

(2.6)
$$
-\partial_t u_{\eta\eta} - \operatorname{tr}[aD^2 u_{\eta\eta}] + D_p H \cdot Du_{\eta\eta} - 2\operatorname{tr}[a_\eta D^2 u_\eta] - \operatorname{tr}[a_{\eta\eta}D^2 u]
$$
$$
+ D_{xx}^2 H\eta \cdot \eta + 2D_{xp}^2 H\eta \cdot Du_\eta + D_{pp}^2 H Du_\eta \cdot Du_\eta = 0.
$$

We now set

$$
U = U(t,x,\xi,\eta) := u_\xi + u_{\eta\eta} - \lambda_0 u \left(1 + \frac{|\xi|^2}{2} + \frac{|\eta|^4}{2} \right)
$$

and compute

$$
D_{(x,\xi,\eta)}^2 U = \begin{pmatrix} D_{xx}^2(u_\xi + u_{\eta\eta}) & D^2 u & 2D_x(D^2 u\eta)^\tau \\ D^2 u & 0 & 0 \\ 2D_x(D^2 u\eta) & 0 & 2D^2 u \end{pmatrix}
$$
$$
- \lambda_0 \begin{pmatrix} D^2 u \left(1 + \frac{|\xi|^2}{2} + \frac{|\eta|^4}{2} \right) & Du \otimes \xi & 2|\eta|^2 Du \otimes \eta \\ \xi \otimes Du & u\operatorname{Id} & 0 \\ 2|\eta|^2 \eta \otimes Du & 0 & 2u(|\eta|^2 \operatorname{Id} + 2\eta \otimes \eta) \end{pmatrix}.
$$

We then set

$$A = A(t, x, \xi, \eta) := \begin{pmatrix} \sigma \\ \sigma_\xi + \sigma_{\eta\eta} \\ \sigma_\eta \end{pmatrix} \begin{pmatrix} \sigma \\ \sigma_\xi + \sigma_{\eta\eta} \\ \sigma_\eta \end{pmatrix}^\tau + \varepsilon \begin{pmatrix} \mathrm{Id}\ 0\ 0 \\ 0\ 0\ 0 \\ 0\ 0\ 0 \end{pmatrix} \geq 0,$$

and add (2.5) and (2.6) and subtract the product of (2.1) with $\lambda_0(1 + |\xi|^2/2 + |\eta|^4/2)$. Using the identities

$$a_\xi = \sigma_\xi \sigma^\tau + \sigma \sigma_\xi^\tau \quad \text{and} \quad a_{\eta\eta} = \sigma_{\eta\eta}\sigma^\tau + 2\sigma_\eta\sigma_\eta^\tau + \sigma\sigma_{\eta\eta}^\tau,$$

we obtain, for a constant $C > 0$ depending on $\|\sigma\|_{C^{1,1}}$ and the bound for u,

$$-\partial_t U - \mathrm{tr}[A D^2_{(x,\xi,\eta)} U] + D_p H \cdot D_x U$$

$$+ D_x H \cdot \xi + D^2_{xx} H \eta \cdot \eta + 2 D^2_{px} H \eta \cdot Du_\eta + D^2_{pp} H Du_\eta \cdot Du_\eta$$

$$+ \lambda_0 \left(1 + \frac{|\xi|^2}{2} + \frac{|\eta|^4}{2} \right)(Du \cdot D_p H - H)$$

$$\leq C(\|Du_t\|_\infty + 1)(|\xi|^2 + |\eta|^4).$$

Using (1.12) and the inequalities

$$1 + \frac{|\xi|^2}{2} + \frac{|\eta|^4}{2} \geq |\xi| + |\eta|^2 \quad \text{and} \quad p \cdot D_p H(x, p) - H(x, p) \geq H(x, 0),$$

this further becomes, for a different constant,

$$(2.7) \qquad -\partial_t U - \mathrm{tr}[A D^2_{(x,\xi,\eta)} U] + D_p H \cdot D_x U \leq C(\|Du_t\|_\infty + 1)(1 + |\xi|^2 + |\eta|^4).$$

Now, for $\mu, \delta > 0$ and some $M(t)$ to be determined, assume for the sake of contradiction that

$$U(t, x, \xi, \eta) - M(t)(1 + |\xi|^2 + |\eta|^4) - \delta|x|^2 - \mu(T - t)$$

attains its maximum at some $(t_0, x_0, \xi_0, \eta_0)$ with $t_0 < T$. By standard arguments, we have $|x_0| \leq C\delta^{-1/2}$ for some constant $C > 0$. In view of (1.11), there exists $\bar{C} > 0$, depending on $\|Du\|_\infty$ (and therefore, a priori, on ε) such that

$$|D_p H_{t_0}(x_0, Du_{t_0}(x_0))| \leq \overline{C}.$$

Plugging into (2.7) and using $A \geq 0$ yields

$$\mu - C\delta - \overline{C}\delta^{1/2} + [-\dot{M}(t_0) - CM - C(\|Du_t\|_\infty + 1)] \left(1 + \frac{|\xi|^2}{2} + \frac{|\eta|^4}{2} \right) \leq 0.$$

A contradiction is reached upon taking δ sufficiently small, depending on μ and $\|Du\|_\infty$, and setting, for arbitrary $\alpha > 0$,

$$M(t) := \alpha e^{C(T-t)} + C \int_t^T e^{C(s-t)}(\|Du_s\|_\infty + 1)ds.$$

Taking first $\delta \to 0$ and then $\mu \to 0$, we deduce that, for all $t \in [0, T]$ and $\xi, \eta \in \mathbb{R}^d$,

$$u_{t,\eta\eta} + u_{t,\xi} - \left[C \int_t^T e^{C(s-t)}(\|Du_s\|_\infty + 1)ds + \lambda_0 u_t + \alpha e^{C(T-t)} \right]$$

$$(2.8) \qquad \cdot \left(1 + \frac{|\xi|^2}{2} + \frac{|\eta|^4}{2} \right)$$

$$\leq \sup_{x,\xi,\eta \in \mathbb{R}^d} \left\{ G_{\eta\eta}(x) + G_\xi(x) - [\lambda_0 G(x) + \alpha] \left(1 + \frac{|\xi|^2}{2} + \frac{|\eta|^4}{2} \right) \right\}.$$

Maximizing the right-hand side of (2.8) in ξ and η yields

$$\sup_{x\in\mathbb{R}^d}\left\{\frac{m_+(D^2G(x))^2+|DG(x)|^2}{2(\alpha+\lambda_0 G(x))}-(\alpha+\lambda_0 G(x))\right\},$$

which is nonpositive if we take

$$\alpha=\sup_{x\in\mathbb{R}^d}\left\{\frac{1}{\sqrt{2}}\left(m_+(D^2G(x))^2+|DG(x)|^2\right)^{1/2}-\lambda_0 G(x)\right\}_+$$

(note that in that case $\alpha+\lambda_0 G(x)\geq 0$ for all $x\in\mathbb{R}^d$). Thus, (2.8) becomes, for all $t\in[0,T]$ and $\xi,\eta\in\mathbb{R}^d$,

$$u_{t,\eta\eta}+u_{t,\xi}-\left[C\int_t^T e^{C(s-t)}(\|Du_s\|_\infty+1)ds+\lambda_0 u_t+\alpha e^{C(T-t)}\right]\left(1+\frac{|\xi|^2}{2}+\frac{|\eta|^4}{2}\right)\leq 0.$$

Maximizing in ξ and η gives, for all $t\in[0,T]$,

$$\frac{m_+(D^2u_t)^2+|Du_t|^2}{2\left[C\int_t^T e^{C(s-t)}(\|Du_s\|_\infty+1)ds+\lambda_0 u_t+\alpha e^{C(T-t)}\right]}$$

$$-C\int_t^T e^{C(s-t)}(\|Du_s\|_\infty+1)ds-\lambda_0 u_t-\alpha e^{C(T-t)}\leq 0.$$

Rearranging terms and using the definition of α yields that the quantity

$$\gamma_t:=\sup_{x\in\mathbb{R}^d}\left\{\left(m_+(D^2u_t(x))^2+|Du_t(x)|^2\right)^{1/2}-\sqrt{2}\lambda_0 u_t(x)\right\}_+$$

satisfies

$$\gamma_t\leq C\int_t^T e^{C(s-t)}(\|Du_s\|_\infty+1)ds+e^{C(T-t)}\gamma_T.$$

From (2.2) and Cauchy-Schwartz inequality, we deduce $\|Du_s\|_\infty\leq\gamma_s+C$, and so Grönwall's inequality gives the desired bound

$$\gamma_t\leq C(T-t)+e^{C(T-t)}\gamma_T.$$

$\square$

3. The backward stochastic HJB equation.

We define a notion of a weak solution $(\widetilde{u},M)$ (Definition 3.1) for the terminal value problem

$$(3.1)\quad\begin{cases}d\widetilde{u}_t=\left[-\operatorname{tr}(\widetilde{a}_t(x)D^2\widetilde{u}_t)+\widetilde{H}_t(x,D\widetilde{u}_t)\right]dt+dM_t=0 & \text{in }(0,T)\times\mathbb{R}^d \quad\text{and}\\ \widetilde{u}_T=\widetilde{G} & \text{on }\mathbb{R}^d.\end{cases}$$

More precisely, the solutions are weak in the PDE sense, and strong in the probabilistic sense, since the unknown $(\widetilde{u},M)$ is required to be adapted with respect to the given underlying filtration.

The concept of solutions is based on a semiconcavity bound in space, which will imply in particular that $D^2\widetilde{u}$ is a locally finite measure. We then prove the existence, uniqueness, and stability of such solutions (Theorems 3.1, 3.2, and 3.3).

We fix a filtered probability space $(\Omega,\mathbb{F},\mathbb{P})$ such that

$$(3.2)\qquad\mathbb{F}=(\mathbb{F}_t)_{t\in[0,T]}\quad\text{is a filtration satisfying the usual conditions,}$$

and we assume that

$$(3.3) \quad \begin{cases} \widetilde{a} : \Omega \times [0,T] \times \mathbb{R}^d \to \mathbb{S}^d \text{ and } \widetilde{H} : \Omega \times [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \text{ are progressively} \\ \quad \text{measurable with respect to } \mathbb{F} \text{ and the Borel topology,} \\ \widetilde{G} : \Omega \times \mathbb{R}^d \to \mathbb{R} \text{ is } \mathbb{F}_T \otimes \mathcal{B}_{\mathbb{R}^d}\text{-measurable, and} \\ \widetilde{a}, \widetilde{G}, \widetilde{H} \text{ satisfy } (1.9), (1.10), (1.11), \text{ and } (1.12) \text{ uniformly over } \Omega. \end{cases}$$

Recall that equations such as (3.1) arise from transforming the HJB equation in the MFG system (1.1) through the relations (1.2) - (1.6), and therefore (3.3) is completely natural (see Remark 1.1). Note that, throughout this section, the explicit dependence of the data on the Wiener process W is not used.

3.1. *The definition of solutions.*   The notion of solution, and, in particular, the condition on $D^2\widetilde{u}$, is motivated by Proposition 2.1, which indicates that $D^2\widetilde{u}$ can be expected to belong to $\mathcal{M}_{\mathrm{loc}} \cap W^{-1,\infty}$.

In what follows, we make use of the quasinorm

$$\|\mathcal{T}\|_{\mathcal{M}_{\mathrm{loc}} \cap W^{-1,\infty}} = \|\mathcal{T}\|_{\mathcal{M}_{\mathrm{loc}}(\mathbb{R}^d)} \vee \|\mathcal{T}\|_{W^{-1,\infty}} .$$

Just as for the quasinorm $\|\cdot\|_{\mathcal{M}_{\mathrm{loc}}}$, this induces a metric on $\mathcal{M}_{\mathrm{loc}} \cap W^{-1,\infty}$. We will say a subset of $\mathcal{M}_{\mathrm{loc}} \cap W^{-1,\infty}$ is bounded if the quasinorm is uniformly bounded on this subset.

We will call a process $X : \Omega \times [0,T] \to \mathcal{M}_{\mathrm{loc}} \cap W^{-1,\infty}$ progressively measurable if $\Omega \times [0,T] \ni (\omega, t) \mapsto \langle X_t(\omega), f \rangle$ is progressively measurable for all $f \in C_c \cup W^{1,1}$, and we will say X is a martingale if $\langle X, f \rangle$ is a martingale for all $f \in C_c \cup W^{1,1}$.

REMARK 3.1.   Much of the analysis to follow involves evaluating the relevant distribution-valued stochastic processes at test functions. It is for this reason that we work with the above notions of measurability/integrability, rather than dealing directly with the (strong) Borel topology of the metric space $\mathcal{M}_{\mathrm{loc}} \cap W^{-1,\infty}$.

We use the following infinite-dimensional generalization of the well-known fact that martingales adapted to right-continuous filtrations have càdlàg versions.

LEMMA 3.1.   *Assume M is a progressively measurable martingale with respect to $\mathbb{F}$ taking values in a bounded subset of $\mathcal{M}_{\mathrm{loc}} \cap W^{-1,\infty}$. Then there exists a martingale $M' : \Omega \times [0,T] \to \mathcal{M}_{\mathrm{loc}} \cap W^{-1,\infty}$ such that*

(a) *For $\mathbb{P}$-a.e. $\omega \in \Omega$,*

$$t \mapsto \langle M_t'(\omega), f \rangle \quad \text{is càdlàg for all } f \in C_c \cup W^{1,1}.$$

(b) *For all $t \in [0,T]$, $M_t(\omega) = M_t'(\omega)$ for $\mathbb{P}$-a.e. $\omega \in \Omega$.*

PROOF.   Let $F \subset C_c \cup W^{1,1}$ be countable and dense in both C_c and $W^{1,1}$. Then, for every $f \in F$, $\langle M, f \rangle$ is a real-valued martingale. Therefore there exist events Ω', $(\Omega_t)_{t \in [0,T]}$ with $\mathbb{P}(\Omega') = 1$ and $\mathbb{P}(\Omega_t) = 1$ for all $t \in [0,T]$, and, for all $f \in F$, a martingale M^f such that

$$M_t^f(\omega) = \langle M_t(\omega), f \rangle \quad \text{for all } t \in [0,T], \omega \in \Omega_t, \text{ and } f \in F,$$

and such that $t \mapsto M_t^f(\omega)$ is càdlàg for all $\omega \in \Omega'$.

By the boundedness of M, there exists $C > 0$ such that, for all $f, g \in F$, $t \in [0,T]$, and $\omega \in \Omega_t$,

$$|M_t^f(\omega) - M_t^g(\omega)| = |\langle M_t(\omega), f - g \rangle| \leq C \|f - g\|_{C(\mathbb{R}^d)} \wedge \|f - g\|_{W^{1,1}},$$

and therefore $f \mapsto M_t^f(\omega)$ extends continuously to an element of $\mathcal{M}_{\mathrm{loc}} \cap W^{-1,\infty}$, which we denote by $M_t'(\omega)$.

For $\omega \in \Omega'$, let $I(\omega) := \{t \in [0,T] : \omega \in \Omega_t\}$. Then, by Fubini's theorem, there exists $\Omega'' \subset \Omega'$ of full probability such that, for all $\omega \in \Omega''$, $|[0,T] \backslash I(\omega)| = 0$, and so, for all $\omega \in \Omega''$ all $t \in I(\omega)$ and $f \in C_c \cup W^{1,1}$,

$$M_t^f(\omega) = \langle M_t'(\omega), f \rangle.$$

Because $\omega \in \Omega'$, $t \mapsto M_t^f(\omega)$ is càdlàg for all $f \in F$, and therefore for all $f \in C_c \cup W^{1,1}$ by the boundedness of M and the density of F. Then, because $I(\omega)$ has full measure, it follows that, for any $t \in [0,T]$ and $\omega \in \Omega''$, $f \mapsto M_t^f(\omega)$ is a continuous linear map $M_t'(\omega)$ on $W^{1,1} \cup C_c$. By definition, M' is a progressively measurable martingale taking values in $\mathcal{M}_{\mathrm{loc}} \cap W^{1,1}$, with the càdlàg property of (a) satisfied, and, for all $t \in [0,T]$ and $f \in W^{1,1} \cup C_c$,

$$\langle M_t'(\omega), f \rangle = \langle M_t(\omega), f \rangle \quad \text{for all } \omega \in \Omega_t \cap \Omega''.$$

Since $\mathbb{P}(\Omega_t \cap \Omega'') = 1$ for all $t \in [0,T]$, (b) is satisfied. $\qquad \square$

DEFINITION 3.1. The pair $(\widetilde{u}, M)$ is a solution of (3.1) if

(a) $[0,T] \times \Omega \ni (t,\omega) \to \widetilde{u}_t \in C(\mathbb{R}^d)$ is a càdlàg process progressively measurable with respect to the filtration $\mathbb{F}$, with $\widetilde{u}_T = \widetilde{G}$,
(b) $M : [0,T] \times \Omega \to \mathcal{M}_{\mathrm{loc}}(\mathbb{R}^d) \cap W^{-1,\infty}(\mathbb{R}^d)$ is a càdlàg martingale with respect to the filtration $\mathbb{F}$,
(c) there exists a deterministic constant $C > 0$ such that, with probability one and for a.e. $t \in [0,T]$,

$$\|D\widetilde{u}_t\|_\infty + \operatorname*{ess\,sup} m_+(D^2\widetilde{u}_t) + \|M_t\|_{\mathcal{M}_{\mathrm{loc}} \cap W^{-1,\infty}} \leq C,$$

and
(d) the following is satisfied in the distributional sense on $\mathbb{R}^d$ and for all $0 \leq t \leq T$:

$$(3.4) \qquad \widetilde{u}_t = \widetilde{G} + \int_t^T \left[\operatorname{tr}\left(\widetilde{a}_r D^2 \widetilde{u}_r \right) - \widetilde{H}_r(\cdot, D\widetilde{u}_r) \right] dr + M_t - M_T.$$

REMARK 3.2. In view of the assumptions on the martingale M and Lemma 2.1(c), the distributional equality (3.4) admits test functions $f \in C_c(\mathbb{R}^d) \cup W^{1,1}(\mathbb{R}^d)$.

The following lemma is a useful tool in the proofs of existence and uniqueness of solutions.

LEMMA 3.2. Assume that $[0,T] \times \Omega \ni (t,\omega) \to u_t \in C(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$ and $M : [0,T] \times \Omega \to \mathcal{M}_{\mathrm{loc}}(\mathbb{R}^d) \cap W^{-1,\infty}(\mathbb{R}^d)$ are $\mathbb{F}$-progressively measurable càdlàg processes, M is a martingale, $\mu \in L^\infty(\Omega \times [0,T], \mathcal{M}_{\mathrm{loc}}(\mathbb{R}^d) \cap W^{-1,\infty}(\mathbb{R}^d))$ is $\mathbb{F}$-progressively measurable, and, for all $0 \leq s \leq t \leq T$, in the sense of distributions on $\mathbb{R}^d$,

$$(3.5) \qquad u_t - u_s = \int_s^t \mu_r dr + M_t - M_s.$$

Let $\phi \in C^2(\mathbb{R}) \cap W^{2,\infty}(\mathbb{R})$ be convex, and set $v_t(x) := \phi(u_t(x))$ for $(t,x) \in [0,T] \times \mathbb{R}^d$. Then, for all nonnegative f such that $f, \frac{\partial f}{\partial t} \in C_c([0,T] \times \mathbb{R}^d)$ or $f \in W^{1,1}([0,T] \times \mathbb{R}^d)$,

$$\mathbb{E} \int_{\mathbb{R}^d} v_t(x) f_t(x) dx - \mathbb{E} \int_{\mathbb{R}^d} v_s(x) f_s(x) dx$$

$$\geq \mathbb{E} \int_s^t \left[\langle \mu_r, \phi'(u_r) f_r \rangle + \int_{\mathbb{R}^d} v_r(x) \frac{\partial f_r(x)}{\partial r} dx \right] dr.$$

The same is true if the equality in (3.5) *is replaced with* $\leq$ *(resp.* $\geq$*) and* ϕ *is non-increasing (resp. non-decreasing).*

PROOF. We prove only the claim when (3.5) holds with equality, because the other two claims are argued similarly. Fix a partition $P = \{s = \tau_0 < \tau_1 < \cdots < \tau_N = t\}$ of $[s, t]$. Then, by the convexity of ϕ and the nonnegativity of f,

$$\int_{\mathbb{R}^d} v_t(x) f_t(x) dx - \int_{\mathbb{R}^d} v_s(x) f_s(x) dx$$

$$= \sum_{n=1}^{N} \left(\int_{\mathbb{R}^d} (v_{\tau_n}(x) - v_{\tau_{n-1}}(x)) f_{\tau_{n-1}}(x) dx + \int_{\mathbb{R}^d} v_{\tau_n}(x)(f_{\tau_n}(x) - f_{\tau_{n-1}}(x)) dx \right)$$

$$\geq \sum_{n=1}^{N} \left(\int_{\mathbb{R}^d} (u_{\tau_n}(x) - u_{\tau_{n-1}}(x)) \phi'(u_{\tau_{n-1}}(x)) f_{\tau_{n-1}}(x) dx \right.$$

$$\left. + \int_{\mathbb{R}^d} v_{\tau_n}(x)(f_{\tau_n}(x) - f_{\tau_{n-1}}(x)) dx \right).$$

For every $n = 1, 2, \ldots, N$, $\phi'(u_{\tau_{n-1}})$ is $\mathbb{F}_{\tau_{n-1}}$-measurable, and so, by the martingale property for M, for $\mathbb{P}$-a.e. $\omega \in \Omega$,

$$\mathbb{E}\left[\langle M_{\tau_n} - M_{\tau_{n-1}}, \phi'(u_{\tau_{n-1}}) f_{\tau_{n-1}} \rangle \mid \mathbb{F}_{\tau_{n-1}} \right](\omega) = 0.$$

The nested property of conditional expectations and (3.5) yield

$$\mathbb{E} \int_{\mathbb{R}^d} (u_{\tau_n}(x) - u_{\tau_{n-1}}(x)) \phi'(u_{\tau_{n-1}}(x)) f_{\tau_{n-1}}(x) dx = \mathbb{E} \int_{\tau_{n-1}}^{\tau_n} \langle \mu_r, \phi'(u_{\tau_{n-1}}) f_{\tau_{n-1}} \rangle dr,$$

and so

$$\mathbb{E} \int_{\mathbb{R}^d} v_t(x) f_t(x) dx - \mathbb{E} \int_{\mathbb{R}^d} v_s(x) f_s(x) dx$$

$$\geq \mathbb{E} \int_s^t \sum_{n=1}^{N} \mathbf{1}_{[\tau_{n-1}, \tau_n)}(r) \left(\langle \mu_r, \phi'(u_{\tau_{n-1}}) f_{\tau_{n-1}} \rangle + \int_{\mathbb{R}^d} v_{\tau_n}(x) \frac{\partial f_r(x)}{\partial r} dx \right) dr.$$

Denote by θ^P the process

$$\theta_r^P := \sum_{n=1}^{N} \mathbf{1}_{[\tau_{n-1}, \tau_n)}(r) \left(\langle \mu_r, \phi'(u_{\tau_{n-1}}) f_{\tau_{n-1}} \rangle + \int_{\mathbb{R}^d} v_{\tau_n}(x) \frac{\partial f_r(x)}{\partial r} dx \right), \quad r \in [s, t].$$

Then $\operatorname{ess\,sup}_{(r,\omega) \in [s,t] \times \Omega} |\theta_r^P|$ is bounded independently of P, and, as $|P| \to 0$, with probability one and for almost every $r \in [s, t]$, θ_r^P converges to

$$\theta_r := \langle \mu_r, \phi'(u_r) f \rangle + \int_{\mathbb{R}^d} v_r(x) \frac{\partial f_r(x)}{\partial r} dx.$$

We conclude upon sending $|P| \to 0$ and appealing to the dominated convergence theorem. $\square$

We will apply the preceding lemma with a particular class of test functions. For $\lambda > 0$ and $\mathcal{K} > 0$, define

$$(3.6) \qquad \psi_{\lambda, \mathcal{K}}(t, x) := \exp\left(-\frac{[(|x| - \mathcal{K}t)_+]^2}{4\lambda(t+1)} \right), \quad (t, x) \in [0, T] \times \mathbb{R}^d.$$

LEMMA 3.3. *For any $\lambda > 0$ and $\mathcal{K} > 0$, $\psi_{\lambda,K}$ belongs to $C^2([0,T] \times \mathbb{R}^d) \cap W^{2,1}([0,T] \times \mathbb{R}^d)$ and satisfies*

$$\frac{\partial \psi_{\lambda,\mathcal{K}}}{\partial t} - \lambda m_+(D^2 \psi_{\lambda,\mathcal{K}}) - \mathcal{K}|D\psi_{\lambda,\mathcal{K}}| \geq 0 \quad in \ [0,T] \times \mathbb{R}^d.$$

PROOF. It is clear that the derivatives of $\psi_{\lambda,\mathcal{K}}$ are continuous and belong to $L^1([0,T] \times \mathbb{R}^d)$. To check the differential inequality, we compute

$$\frac{\partial \psi_{\lambda,\mathcal{K}}}{\partial t} = \exp\left(-\frac{[(|x| - \mathcal{K}t)_+]^2}{4\lambda(t+1)}\right)\left[\frac{[(|x| - \mathcal{K}t)_+]^2}{4\lambda(t+1)^2} + \frac{\mathcal{K}(|x| - \mathcal{K}t)_+}{2\lambda(t+1)}\right],$$

$$D\psi_{\lambda,\mathcal{K}}(t,x) = -\exp\left(-\frac{[(|x| - \mathcal{K}t)_+]^2}{4\lambda(t+1)}\right)\frac{(|x| - \mathcal{K}t)_+}{2\lambda(t+1)}\hat{x},$$

and

$$D^2\psi_{\lambda,\mathcal{K}}(t,x) = \mathrm{sgn}_+(|x| - \mathcal{K}t)\exp\left(-\frac{[(|x| - \mathcal{K}t)_+]^2}{4\lambda(t+1)}\right)$$

$$\cdot\left[-\frac{|x| - \mathcal{K}t}{2|x|\lambda(t+1)}\,\mathrm{Id} + \left(\frac{(|x| - \mathcal{K}t)^2}{4\lambda^2(t+1)^2} + \frac{|x| - \mathcal{K}t}{2|x|\lambda(t+1)} - \frac{1}{2\lambda(t+1)}\right)\hat{x} \otimes \hat{x}\right].$$

The identity $m_+(A\hat{x} \otimes \hat{x} + B\,\mathrm{Id}) = (B + A_+)_+$ for all $A, B \in \mathbb{R}$ and $x \in \mathbb{R}^d$ yields

$$m_+(D^2\psi_{\lambda,\mathcal{K}}(t,x)) = \mathrm{sgn}_+(|x| - \mathcal{K}t)\exp\left(-\frac{[(|x| - \mathcal{K}t)_+]^2}{4\lambda(t+1)}\right)$$

$$\cdot\left[-\frac{|x| - \mathcal{K}t}{2|x|\lambda(t+1)} + \left(\frac{(|x| - \mathcal{K}t)^2}{4\lambda^2(t+1)^2} + \frac{|x| - \mathcal{K}t}{2|x|\lambda(t+1)} - \frac{1}{2\lambda(t+1)}\right)_+\right]_+$$

$$= \mathrm{sgn}_+(|x| - \mathcal{K}t)\exp\left(-\frac{[(|x| - \mathcal{K}t)_+]^2}{4\lambda(t+1)}\right)\left(\frac{(|x| - \mathcal{K}t)^2}{4\lambda^2(t+1)^2} - \frac{1}{2\lambda(t+1)}\right)_+.$$

We conclude that

$$\frac{\partial \psi_{\lambda,\mathcal{K}}}{\partial t} - \lambda m_+(D^2\psi_{\lambda,\mathcal{K}}) - \mathcal{K}|D\psi_{\lambda,\mathcal{K}}|$$

$$\geq \mathrm{sgn}_+(|x| - \mathcal{K}t)\exp\left(-\frac{[(|x| - \mathcal{K}t)_+]^2}{4\lambda(t+1)}\right)\left[\frac{(|x| - \mathcal{K}t)^2}{4\lambda(t+1)^2} - \left(\frac{(|x| - \mathcal{K}t)^2}{4\lambda(t+1)^2} - \frac{1}{2(t+1)}\right)_+\right.$$

$$\left.+ \frac{\mathcal{K}(|x| - \mathcal{K}t)}{2\lambda(t+1)} - \frac{\mathcal{K}(|x| - \mathcal{K}t)}{2\lambda(t+1)}\right] \geq 0.$$

$\square$

3.2. *Existence.* Using a discretization procedure, we prove here the existence of a solution of (3.1).

THEOREM 3.1. *Assume (3.2) and (3.3). Then there exists a solution $(\tilde{u}, M)$ of (3.1) in the sense of Definition 3.1.*

Fix $N \in \mathbb{N}$ and set

(3.7) $$t_n^N := \frac{nT}{2^N} \quad \text{for } n = 0, 1, 2, \ldots, 2^N.$$

Define the time-discretized diffusion and Hamiltonian by

$$(3.8) \qquad \widetilde{a}_t^N := \widetilde{a}_{t_n^N} \quad \text{for } t \in \left[t_{n-1}^N, t_n^N\right), \quad n = 1, 2, \ldots, 2^N$$

and

$$(3.9) \qquad \widetilde{H}_t^N := \widetilde{H}_{t_n^N} \quad \text{for } t \in \left[t_{n-1}^N, t_n^N\right), \quad n = 1, 2, \ldots, 2^N,$$

as well as the right-continuous, piecewise-constant-in-time filtration $\mathbb{F}^N = (\mathbb{F}_t^N)_{t \in [0,T]}$ by

$$(3.10) \qquad \mathbb{F}_t^N := \mathbb{F}_{t_n^N} \quad \text{for } t \in \left[t_{n-1}^N, t_n^N\right), n = 1, 2, \ldots, 2^N, \quad \mathbb{F}_T^N = \mathbb{F}_T.$$

Observe that, in view of the right-continuity of $\mathbb{F}$ implied by (3.2),

$$(3.11) \qquad \mathbb{F}^{N+1} \subset \mathbb{F}^N \quad \text{for all } N \text{ and } \mathbb{F}_t = \bigcap_{N \in \mathbb{N}} \mathbb{F}_t^N \quad \text{for all } t \in [0,T].$$

We now introduce a càdlàg-in-time function $\widetilde{u}^N : [0,T] \times \mathbb{R}^d \times \Omega \to \mathbb{R}$, with jumps occurring at the partition points (3.7), as follows. We first set $\widetilde{u}_T^N := \widetilde{G}$ in $\mathbb{R}^d$. Then, for $n = 1, 2, \ldots, 2^N$, given $\widetilde{u}_{t_n^N}^N$, define $\widetilde{u}^N$ in $\left[t_{n-1}^N, t_n^N\right)$ as the solution of the terminal value problem

$$(3.12) \qquad \begin{cases} -\dfrac{\partial \widetilde{u}_t^N}{\partial t} - \text{tr}[\widetilde{a}_t^N D^2 \widetilde{u}_t^N] + \widetilde{H}_t^N(x, D\widetilde{u}_t^N) = 0 \quad \text{in } \left[t_{n-1}^N, t_n^N\right) \times \mathbb{R}^d \quad \text{and} \\[2mm] \widetilde{u}_{t_n^N, -}^N = \mathbb{E}\left[\widetilde{u}_{t_n^N}^N \,\Big|\, \mathbb{F}_{t_n^N}\right] \quad \text{in } \mathbb{R}^d. \end{cases}$$

In other words, the jump discontinuity of $\widetilde{u}^N$ at t_n^N is of size

$$\Delta M_{t_n^N}^N := \widetilde{u}_{t_n^N}^N - \mathbb{E}\left[\widetilde{u}_{t_n^N}^N \,\Big|\, \mathbb{F}_{t_{n-1}^N}^N\right], \quad n = 1, 2, \ldots, 2^N - 1.$$

Finally, define

$$(3.13) \qquad M_t^N := \sum_{n: t_n^N \leq t} \Delta M_{t_n^N}^N \quad \text{for } t \in [0,T].$$

LEMMA 3.4.   *The processes constructed above satisfy the following:*

(a) *The process $(\widetilde{u}_t^N)_{t \in [0,T]}$ is $\mathbb{F}^N$-adapted, and the process $(M_t^N)_{t \in [0,T]}$ is a $\mathbb{F}^N$-martingale.*
(b) *There exists a constant $C > 0$ depending only on T and the data (through the assumptions listed in the statement of the theorem) such that*

$$\sup_{N \in \mathbb{N}} \left(\left\|\widetilde{u}^N\right\|_{\infty, [0,T] \times \Omega \times \mathbb{R}^d} + \left\|D\widetilde{u}^N\right\|_{\infty, [0,T] \times \Omega \times \mathbb{R}^d} + \operatorname*{ess\,sup}_{[0,T] \times \Omega \times \mathbb{R}^d} m_+(D^2\widetilde{u}^N) \right) \leq C.$$

(c) *For all $t \in [0,T]$ and in the sense of distributions in $\mathbb{R}^d$,*

$$\widetilde{u}_t^N = \widetilde{G} + \int_t^T \left(\text{tr}[\widetilde{a}_s^N D^2 \widetilde{u}_s^N] - \widetilde{H}_s^N(\cdot, D\widetilde{u}_s^N) \right) ds - M_T^N + M_t^N.$$

PROOF.   We prove only (b), and then properties (a) and (c) are easily checked.

Fix $n = 0, 1, \ldots, 2^N - 1$. Combining the bound $\left\|\widetilde{u}_{t_{n+1}^N, -}^N\right\|_\infty \leq \left\|\widetilde{u}_{t_{n+1}^N}^N\right\|_\infty$ with (2.2) from Proposition 2.1 yields, for some $C > 0$ as in the statement of the lemma,

$$\left\|\widetilde{u}_t^N\right\|_\infty \leq \left\|\widetilde{u}_{t_{n+1}^N}^N\right\|_\infty + C(t_{n+1} - t) \quad \text{for all } t \in [t_n^N, t_{n+1}^N).$$

18

We then see inductively that, for all $t \in [0, T]$,

$$\left\| \widetilde{u}_t^N \right\|_\infty \leq \|G\|_\infty + C(T - t).$$

A similar argument, appealing to (2.3) from Proposition 2.1, gives, for all $t \in [0, T]$ and almost everywhere in $\mathbb{R}^d \times \Omega$,

$$\left(m_+ \left(\frac{dD^2\widetilde{u}_t^N}{d\mathcal{L}} \right)^2 + |D\widetilde{u}_t^N|^2 \right)^{1/2} - \sqrt{2}\lambda_0 \widetilde{u}_t^N$$

$$\leq \frac{T}{2^N} \sum_{k=0}^{2^N-1} e^{Ck/2^N} + e^{CT} \operatorname*{ess\,sup}_{\mathbb{R}^d \times \Omega} \left\{ \left(m_+ \left(\frac{dD^2G}{d\mathcal{L}} \right)^2 + |DG|^2 \right)^{1/2} - \sqrt{2}\lambda_0 G \right\}_+ .$$

Combining the above with Lemma 2.1(b), the bound for $\|\widetilde{u}\|_\infty$, and

$$\frac{T}{2^N} \sum_{k=0}^{2^N-1} e^{CkT/2^N} = \frac{T(e^{CT} - 1)}{2^N(e^{CT/2^N} - 1)} \leq \frac{e^{CT} - 1}{C}$$

shows that

$$\sup_{N \in \mathbb{N}} \operatorname*{ess\,sup}_{[0,T] \times \mathbb{R}^d \times \Omega} m_+ \left(\frac{dD^2\widetilde{u}_t^N}{d\mathcal{L}} \right) < \infty,$$

so that, by Lemma 2.1(a) and (b), $\widetilde{u}^N$ is uniformly semiconcave and Lipschitz. □

The proof of Theorem 3.1, as well as for the stability estimates to follow, involve bounding the difference between uniformly Lipschitz solutions in certain Lebesgue spaces. The following elementary lemma allows to translate the last bound to an estimate in the local-uniform topology.

LEMMA 3.5. *Let $1 \leq p < \infty$. Then there exists $C = C_{d,p} > 0$ such that, for all $w \in W^{1,\infty}(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$,*

$$\|w\|_\infty \leq C \|Dw\|_\infty^{\frac{d}{d+p}} \|w\|_p^{\frac{p}{d+p}} .$$

PROOF. Fix $x \in \mathbb{R}^d$. The function $\mathcal{C}_x(y) := (|w(x)| - \|Dw\|_\infty |y - x|)_+$ lies below the graph of $|w|$, which implies that $\|\mathcal{C}_x\|_p \leq \|w\|_p$. For some $c = c_{d,p} > 0$, we compute $\|\mathcal{C}_x\|_p = c|w(x)|^{d+p} \|Dw\|_\infty^{-d}$, and the result follows upon rearranging terms and taking the supremum over $x \in \mathbb{R}^d$. □

PROOF OF THEOREM 3.1. Fix $K, N \in \mathbb{N}$ with $N < K$, let $\phi : \mathbb{R} \to \mathbb{R}$ be smooth, non-negative, nonincreasing, and convex, and set

$$w_t^{N,K}(x) := \phi\left(\widetilde{u}_t^N(x) - \widetilde{u}_t^K(x) \right) \quad \text{for } (t, x) \in [0, T] \times \mathbb{R}^d.$$

Let $f : [0, T] \times \mathbb{R}^d \to \mathbb{R}$ be smooth and nonnegative, with sufficient decay at infinity (to be determined below). For $0 \le t < t + h \le T$,

$$\int_{\mathbb{R}^d} \left(w_{t+h}^{N,K}(x) f_{t+h}(x) - w_t^{N,K}(x) f_t(x) \right) dx$$

$$= \int_t^{t+h} \left[- \langle \mathrm{tr}[\tilde{a}_s^N D^2 \tilde{u}_s^N - \tilde{a}_s^K D^2 \tilde{u}_s^K], f_s \phi'(\tilde{u}_s^N - \tilde{u}_s^K) \rangle \right.$$

(3.14)
$$+ \int_{\mathbb{R}^d} \left([\tilde{H}_s^N(x, D\tilde{u}_s^N(x)) - \tilde{H}_s^K(x, D\tilde{u}_s^K(x))] f_s(x) \phi'(\tilde{u}_s^N(x) - \tilde{u}_s^K(x)) \right.$$

$$\left. \left. + w_s^{N,K}(x) \frac{\partial f_s(x)}{\partial s} \right) dx \right] ds$$

$$+ \sum_{t_n^N \in (t, t+h]} \int_{\mathbb{R}^d} \Delta w_{t_n^N}^{N,K}(x) f_{t_n^N}(x) dx + \sum_{t_n^K \in (t, t+h]} \int_{\mathbb{R}^d} \Delta w_{t_n^K}^{N,K}(x) f_{t_n^K}(x) dx,$$

where

$$\Delta w_{t_n^N}^{N,K} := \phi\left(\tilde{u}_{t_n^N}^N - \tilde{u}_{t_n^N}^K \right) - \phi\left(\tilde{u}_{t_n^N, -}^N - \tilde{u}_{t_n^N, -}^K \right)$$

and

$$\Delta w_{t_n^K}^{N,K} := \phi\left(\tilde{u}_{t_n^K}^N - \tilde{u}_{t_n^K}^K \right) - \phi\left(\tilde{u}_{t_n^K, -}^N - \tilde{u}_{t_n^K, -}^K \right).$$

The convexity of ϕ implies that

$$\Delta w_{t_n^N}^{N,K} \ge \phi'\left(\tilde{u}_{t_n^N, -}^N - \tilde{u}_{t_n^N, -}^K \right) \left(\Delta M_{t_n^N}^N - \Delta M_{t_n^N}^K \right)$$

and

$$\Delta w_{t_n^K}^{N,K} \ge \phi'\left(\tilde{u}_{t_n^K, -}^N - \tilde{u}_{t_n^K, -}^K \right) \left(\Delta M_{t_n^K}^N - \Delta M_{t_n^K}^K \right).$$

By definition, $\phi'(\tilde{u}_{t_n^N, -}^N - \tilde{u}_{t_n^N, -}^K)$ is $\mathbb{F}_{t_n^N}$-measurable, and so

$$\mathbb{E}\left[\Delta w_{t_n^N}^{N,K} \,\Big|\, \mathbb{F}_{t_n^N} \right] \ge 0.$$

If t_n^K is not equal to t_m^N for some $m = 1, 2, \dots, 2^N - 1$, then $\Delta M_{t_n^K}^N = M_{t_n^K}^N - M_{t_n^K, -}^N = 0$. Otherwise, $\phi'(\tilde{u}_{t_n^K, -}^N - \tilde{u}_{t_n^K, -}^K)$ is $\mathbb{F}_{t_n^K}$-measurable. In any case, we have

$$\mathbb{E}\left[\Delta w_{t_n^K}^{N,K} \,\Big|\, \mathbb{F}_{t_n^K} \right] \ge 0.$$

Upon taking the expectation of (3.14), the nested property of conditional expectation, the convexity of ϕ, and the nonnegativity of $\tilde{a}^N$ imply that

$$\mathbb{E} \int_{\mathbb{R}^d} \left(w_{t+h}^{N,K}(x) f_{t+h}(x) - w_t^{N,K}(x) f_t(x) \right) dx$$

$$\ge \int_t^{t+h} \mathbb{E}\left[- \langle \mathrm{tr}[\tilde{a}_s^N D^2 \tilde{u}_s^N - \tilde{a}_s^K D^2 \tilde{u}_s^K], f_s \phi'(\tilde{u}_s^N - \tilde{u}_s^K) \rangle \right.$$

$$+ \int_{\mathbb{R}^d} \left([\tilde{H}_s^N(x, D\tilde{u}_s^N(x)) - \tilde{H}_s^K(x, D\tilde{u}_s^K(x))] f_s(x) \phi'(\tilde{u}_s^N(x) - \tilde{u}_s^K(x)) \right.$$

$$+ w_s^{N,K}(x) \frac{\partial f_s(x)}{\partial s}\bigg) dx \bigg] ds$$

$$\geq \int_t^{t+h} \mathbb{E}\bigg[\Big\langle -\operatorname{tr}[\widetilde{a}_s^N D^2 \widetilde{w}_s^{N,K}] + \zeta_s^{N,K}, f_s \Big\rangle$$

$$+ \int_{\mathbb{R}^d} \bigg(b_s^{N,K}(x) \cdot D w_s^{N,K}(x) f_s(x) + w_s^{N,K}(x) \frac{\partial f_s(x)}{\partial s} \bigg) dx \bigg] ds,$$

where

$$b_s^{N,K} := \int_0^1 D_p \widetilde{H}_s^N(\cdot, \tau D\widetilde{u}_s^N + (1-\tau) D\widetilde{u}_s^K) d\tau$$

and

$$\zeta_s^{N,K} := \phi'(\widetilde{u}_s^N - \widetilde{u}_s^K)\Big[-\operatorname{tr}(\widetilde{a}_s^N - \widetilde{a}_s^K) D^2 \widetilde{u}_s^K + \widetilde{H}_s^N(\cdot, D\widetilde{u}_s^K) - \widetilde{H}_s^K(\cdot, D\widetilde{u}_s^K) \Big].$$

For $\delta > 0$, let ρ_δ be as in (1.15) and define

$$b_s^{N,K,\delta} := \int_0^1 D_p \widetilde{H}_s^N \left(\cdot, \tau D(\rho_\delta * \widetilde{u}_s^N) + (1-\tau) D(\rho_\delta * \widetilde{u}_s^K) \right) d\tau$$

and

$$\zeta_s^{N,K,\delta} := \zeta_s^{N,K} + \left(b_s^{N,K} - b_s^{N,K,\delta} \right) \cdot D w_s^{N,K}.$$

Then

$$\sup_{t \in [0,T]} \left\| b_t^{N,K,\delta} \right\|_\infty \leq \|b\|_\infty \quad \text{and}$$

$$\lim_{\delta \to 0} b^{N,K,\delta} = b^{N,K} \quad \text{almost everywhere in } [0,T] \times \mathbb{R}^d \times \Omega,$$

and, in view of the convexity of $\widetilde{H}$, the uniform semiconcavity bounds for $\widetilde{u}^N$ and $\widetilde{u}^K$ implied by Lemma 3.4(b), and (1.9),

$$\sup_{t \in [0,T]} \sup_{x \in \mathbb{R}^d} \sup_{\delta > 0} \sup_{N,K} \left(\operatorname{div} b_t^{N,K,\delta}(x) + \operatorname{div} \operatorname{div} \widetilde{a}_t^N(x) \right) < \infty.$$

We next write

$$\mathbb{E} \int_{\mathbb{R}^d} \left(w_{t+h}^{N,K}(x) f_{t+h}(x) - w_t^{N,K}(x) f_t(x) \right) dx$$

$$\geq \int_t^{t+h} \mathbb{E}\bigg[\Big\langle -\operatorname{tr}[\widetilde{a}_s^N D^2 w_s^{N,K}] + \zeta_s^{N,K,\delta}, f_s \Big\rangle$$

$$+ \int_{\mathbb{R}^d} \bigg(b_s^{N,K,\delta}(x) \cdot D w_s^{N,K}(x) f_s(x) + w_s^{N,K}(x) \frac{\partial f_s(x)}{\partial s} \bigg) dx \bigg] ds$$

$$= \int_t^{t+h} \mathbb{E}\bigg[\int_{\mathbb{R}^d} w_s^{N,K}(x) \bigg(\frac{\partial f_s(x)}{\partial s} - \operatorname{tr}[D^2(\widetilde{a}_s^N(x) f_s(x))]$$

$$- b_s^{N,K,\delta}(x) \cdot D f_s(x) - (\operatorname{div} b_s^{N,K,\delta}(x)) f_s(x) \bigg) dx + \Big\langle \zeta_s^{N,K,\delta}, f_s \Big\rangle \bigg] ds$$

$$= \int_t^{t+h} \mathbb{E}\Bigg[\int_{\mathbb{R}^d} w_s^{N,K}(x)\Bigg(\frac{\partial f_s(x)}{\partial s} - \mathrm{tr}[\widetilde{a}_s^N(x)D^2 f_s(x)]$$

$$- \Big(b_s^{N,K,\delta}(x) + 2\,\mathrm{div}\,\widetilde{a}_s^N(x) \Big) \cdot Df_s(x) - \Big(\mathrm{div}\, b^{N,K,\delta}(x) + \mathrm{div}\,\mathrm{div}\,\widetilde{a}_s^N(x) \Big) f_s(x) \Bigg) dx$$

$$+ \Big\langle \zeta_s^{N,K,\delta}, f_s \Big\rangle \Bigg] ds.$$

For suitable λ and $\mathcal{K}$ depending only on $\|\widetilde{a}\|_{C^{1,1}}$, the constant C from Lemma 3.4, and the local bound for $D_p\widetilde{H}$ and $D_x\widetilde{H}$ from (1.11), we take $f = \psi_{\lambda,\mathcal{K}}$ from Lemma 3.3 and set

$$\gamma_t^{N,K} := \mathbb{E}\int_{\mathbb{R}^d} w_s^{N,K}(x)\psi_{\lambda,\mathcal{K},s}(x)dx.$$

Then

$$\dot{\gamma}_t^{N,K} \geq -C\gamma_t^{N,K} + \varepsilon_t^{N,K,\delta}, \quad \text{where} \quad \varepsilon_t^{N,K,\delta} := \mathbb{E}\Big\langle \zeta_t^{N,K,\delta}, \psi_{\lambda,\mathcal{K},t} \Big\rangle.$$

The uniform Lipschitz bound for $w^{N,K}$, the fact that $\psi_{\lambda,\mathcal{K}} \in W^{2,1}$, and the dominated convergence theorem yield, upon sending $\delta \to 0$,

$$\dot{\gamma}_t^{N,K} \geq -C\gamma_t^{N,K} + \varepsilon_t^{N,K}, \quad \text{where} \quad \varepsilon_t^{N,K} := \mathbb{E}\Big\langle \zeta_t^{N,K}, \psi_{\lambda,\mathcal{K},t} \Big\rangle.$$

We have $\gamma_T^{N,K} = 0$, and so Grönwall's inequality and the continuity of $\widetilde{H}$ and $\widetilde{a}$ in time give, for all $t \in [0,T]$ and some sequence ω_N satisfying $\lim_{N\to\infty}\omega_N = 0$,

$$\gamma_t^{N,K} \leq \widetilde{C}\int_t^T \varepsilon_s^{N,K}ds \leq \omega_N.$$

We then take $\phi(r) := r_-$ (which can be justified with a smooth approximation) and, upon switching the roles of $\widetilde{u}^N$ and $\widetilde{u}^K$, conclude that

$$\sup_{t\in[0,T]} \mathbb{E}\int_{\mathbb{R}^d} |\widetilde{u}_t^N(x) - \widetilde{u}_t^K(x)|\psi_{\lambda,\mathcal{K},t}(x)dx \leq \omega_N.$$

Lemma 3.5 with $p = 1$ and the uniform Lipschitz bounds for $\widetilde{u}^N$, $\widetilde{u}^K$, and $\psi_{\lambda,\mathcal{K}}$ then give a constant, independent of N and K, such that

$$(3.15) \qquad \sup_{t\in[0,T]} \mathbb{E}\sup_{x\in\mathbb{R}^d} \psi_{\lambda,\mathcal{K},t}(x)^{d+1}|\widetilde{u}_t^N(x) - \widetilde{u}_t^K(x)|^{d+1} \leq C\omega_N.$$

It follows that, as $N \to \infty$, $\widetilde{u}^N$ converges, in $L^\infty([0,T], L^{d+1}(\Omega, C(B_R)))$ for all $R > 0$, to some limit $\widetilde{u}$. Upon extracting a sub-sequence, we deduce that, as $N \to \infty$, $\mathbb{P}$-almost surely and for a.e. $t \in [0,T]$, $\widetilde{u}_t^N$ converges locally uniformly to $\widetilde{u}_t$, and, moreover, because of the uniform semiconcavity bound and Lemma 2.1(d), $D\widetilde{u}_t^N$ converges almost everywhere to $D\widetilde{u}_t$. In view of (3.11), the process $\widetilde{u}$ is progressively measurable with respect to $\mathbb{F}$.

For any $t \in [0,T]$, in the sense of distributions,

$$M_t^N = \widetilde{u}_t^N - \widetilde{u}_0^N - \int_0^t \Big[\mathrm{tr}[\widetilde{a}_s^N D^2\widetilde{u}_s^N] - \widetilde{H}_s^N(\cdot, D\widetilde{u}_s^N) \Big] ds.$$

As $N \to \infty$, for a.e. $t \in [0,T]$ and $\mathbb{P}$-a.s., the right-hand side converges in the distributional sense to

$$(3.16) \quad M_t := \widetilde{u}_t - \widetilde{u}_0 - \int_0^t \Big[\mathrm{tr}[\widetilde{a}_s D^2\widetilde{u}_s] - \widetilde{H}_s(\cdot, D\widetilde{u}_s) \Big] ds \in \mathcal{M}_{\mathrm{loc}}(\mathbb{R}^d) \cap W^{-1,\infty}(\mathbb{R}^d).$$

The process M is progressively measurable with respect to $\mathbb{F}$ because $\widetilde{u}$ is.

For fixed $f \in C_c(\mathbb{R}^d) \cup W^{1,1}(\mathbb{R}^d)$, $(\langle M_t^N, f \rangle)_{t \in [0,T]}$ is a martingale with respect to the filtration $\mathbb{F}^N$. Fix $0 \leq s < t \leq T$ and $\mathcal{A} \in \mathbb{F}_s$. For any $N \in \mathbb{N}$, $\mathbb{F}_s \subset \mathbb{F}_s^N$, and so

$$\mathbb{E}\left[\langle M_t^N - M_s^N, f \rangle \mathbf{1}_{\mathcal{A}}\right] = 0.$$

Sending $N \to \infty$ and appealing to Lemma 3.4(b) and the dominated convergence theorem yields $\mathbb{E}\left[\langle M_t - M_s, f \rangle \mathbf{1}_{\mathcal{A}}\right] = 0$. We conclude that M is a progressively measurable martingale, and then, by Lemma 3.1, M has a version such that, with probability one, $\langle M', f \rangle$ is càdlàg for all $f \in C_c \cup W^{1,1}$.

For $t \in [0,T]$, define

$$\widetilde{u}'_t := \widetilde{G} + M'_t - M'_T - \int_t^T \left[\operatorname{tr}[\widetilde{a}_s D^2 \widetilde{u}_s] - \widetilde{H}_s(\cdot, D\widetilde{u}_s)\right] ds.$$

Then, for all $t \in [0,T]$, there exists $\Omega_t \in \mathbb{F}_t$ with $\mathbb{P}(\Omega_t) = 1$ such that $\widetilde{u}'_t(\omega) = \widetilde{u}_t(\omega)$ for all $\omega \in \Omega_t$, and so the same boundedness, Lipschitz regularity, and semiconcavity properties are satisfied by $\widetilde{u}'$. It follows from Fubini's theorem, that for $\mathbb{P}$-a.e. $\omega \in \Omega$, M' is right-continuous and

$$|\{t \in [0,T] : \omega \notin \Omega_t\}| = 0.$$

Therefore, in the integration over $s \in [t,T]$ above, we may replace $D\widetilde{u}_s$ and $D^2\widetilde{u}_s$ with respectively $D\widetilde{u}'_s$ and $D^2\widetilde{u}'_s$ without affecting the value of the integral. For the rest of the proof, we thus work with the versions of M and $\widetilde{u}$ for which (3.16) holds and $\langle M, f \rangle$ càdlàg for all $f \in C_c \cup W^{1,1}$.

We next note that, by the semiconcavity and Lipschitz bounds for $\widetilde{u}$,

$$t \mapsto \int_0^t \left[\operatorname{tr}[\widetilde{a}_s D^2 \widetilde{u}_s] - \widetilde{H}_s(\cdot, D\widetilde{u}_s)\right] ds$$

is Lipschitz continuous with respect to the strong topology of $\mathcal{M}_{\mathrm{loc}} \cap W^{-1,\infty}$, and therefore, for all $f \in W^{1,1} \cup C_c$, $t \mapsto \langle \widetilde{u}'_t, f \rangle$ is càdlàg. Exploiting the uniform Lipschitz continuity of $\widetilde{u}$ in space, it follows that, in fact, $t \mapsto \widetilde{u}_t$ is càdlàg with respect to local-uniform convergence in $C(\mathbb{R}^d)$. We conclude that $t \mapsto M_t$ is càdlàg with respect to the strong topology of $\mathcal{M}_{\mathrm{loc}} \cap W^{-1,\infty}$ induced by the quasinorm $\|\cdot\|_{\mathcal{M}_{\mathrm{loc}} \cap W^{-1,\infty}}$. $\quad\square$

3.3. *Sub/supersolutions and a comparison principle.*

We relax the notion of solutions of (3.1) with the following definition.

DEFINITION 3.2. The pair $(\widetilde{u}, M)$ is a sub- (resp. super-) solution of (3.1) if (a), (b), and (c) from Definition 3.1 are satisfied, and if, with probability one and for $0 \leq t_1 \leq t_2 \leq T$, in the sense of distributions on $\mathbb{R}^d$,

$$(3.17) \qquad \widetilde{u}_{t_1} \leq \widetilde{u}_{t_2} + \int_{t_1}^{t_2} \left[\operatorname{tr}(\widetilde{a}_s D^2 \widetilde{u}_s) - \widetilde{H}_s(\cdot, D\widetilde{u}_s)\right] ds - M_{t_2} + M_{t_1}$$

(resp.

$$(3.18) \qquad \widetilde{u}_{t_1} \geq \widetilde{u}_{t_2} + \int_{t_1}^{t_2} \left[\operatorname{tr}(\widetilde{a}_s D^2 \widetilde{u}_s) - \widetilde{H}_s(\cdot, D\widetilde{u}_s)\right] ds - M_{t_2} + M_{t_1}\right).$$

Uniqueness of solutions for (3.1) then follows from the following theorem and the fact that solutions are both sub- and super-solutions.

THEOREM 3.2.   *Assume* (3.2), $(\widetilde{a}, \widetilde{H}^1, \widetilde{G}^1)$ *and* $(\widetilde{a}, \widetilde{H}^2, \widetilde{G}^2)$ *satisfy* (3.3), $\widetilde{H}^1 \leq \widetilde{H}^2$, *and* $\widetilde{G}^1 \geq \widetilde{G}^2$, *and let* $(\widetilde{u}^1, M^1)$ *and* $(\widetilde{u}^2, M^2)$ *be respectively a super- and sub-solution of* (3.1) *corresponding to* $(\widetilde{H}^1, \widetilde{G}^1)$ *and* $(\widetilde{H}^2, \widetilde{G}^2)$. *Then* $\widetilde{u}^1 \geq \widetilde{u}^2$.

PROOF.   Let $v = \widetilde{u}^1 - \widetilde{u}^2$ and $M = M^1 - M^2$, and, for $(t, x) \in [0, T] \times \mathbb{R}^d$, define

$$b_t(x) := \int_0^1 D\widetilde{H}_t^1\left(x, \tau D\widetilde{u}_t^1(x) + (1 - \tau)D\widetilde{u}_t^2(x)\right) d\tau.$$

Then

$$-dv_t \geq \left[\mathrm{tr}(\widetilde{a}_t D^2 v_t) - b_t \cdot Dv_t\right] ds - dM_t \quad \text{in } [0, T].$$

Let $\phi : \mathbb{R} \to \mathbb{R}$ be C^2, non-negative, non-increasing, and convex, and set $w = \phi(v)$. Then, by Lemma 3.2, for any nonnegative $f \in W^{2,1}([0,T] \times \mathbb{R}^d) \cap C^2([0,T] \times \mathbb{R}^d)$ and $0 \leq s < t \leq T$,

$$\mathbb{E} \int_{\mathbb{R}^d} w_t(x) f_t(x) dx - \mathbb{E} \int_{\mathbb{R}^d} w_s(x) f_s(x) dx$$

$$\geq \int_s^t \mathbb{E}\left[-\left\langle \mathrm{tr}[\widetilde{a}_r D^2 v_r], \phi'(v_r) f_r \right\rangle + \int_{\mathbb{R}^d} \left(\phi'(v_r(x)) b_r(x) \cdot Dv_r(x) f_r(x) \right.\right.$$

(3.19)

$$\left.\left. + w_r(x) \frac{\partial f_r(x)}{\partial r} \right) dx \right] dr$$

$$\geq \int_s^t \mathbb{E}\left[-\left\langle \mathrm{tr}[\widetilde{a}_r D^2 w_r], f_r \right\rangle + \int_{\mathbb{R}^d} \left(b_r(x) \cdot Dw_r(x) f_r(x) \right.\right.$$

$$\left.\left. + w_r(x) \frac{\partial f_r(x)}{\partial r} \right) dx \right] dr.$$

Fix $\delta > 0$, let ρ_δ be as in (1.15), and, for $t \in [0, T]$, set

$$b_t^\delta := \int_0^1 D\widetilde{H}_t^1\left(\cdot, \tau D(\rho_\delta * \widetilde{u}_t^1) + (1 - \tau)D(\rho_\delta * \widetilde{u}_t^2) \right) d\tau.$$

Then, as $\delta \to 0$, $b^\delta \to b$ almost everywhere in $[0, T] \times \mathbb{R}^d \times \Omega$. Moreover, because $\widetilde{H}^1$ is convex and u^1 and u^2 are globally Lipschitz and semiconcave, we have, for some $C > 0$ independent of δ,

$$\left\| b^\delta \right\|_\infty \leq C \quad \text{and} \quad \mathrm{div}\, b^\delta \leq C.$$

For $r \in [0, T]$, set $\zeta_r^\delta := -(b_r - b_r^\delta) \cdot Dw_r$. Continuing (3.19), we have

$$\mathbb{E} \int_{\mathbb{R}^d} w_t(x) f_t(x) dx - \mathbb{E} \int_{\mathbb{R}^d} w_s(x) f_s(x) dx$$

$$\geq \int_s^t \mathbb{E}\left[-\left\langle \mathrm{tr}[\widetilde{a}_r D^2 w_r] + \zeta_r^\delta, f_r \right\rangle + \int_{\mathbb{R}^d} \left(\mathrm{div}(b_r^\delta(x) w_r(x)) f_r(x) \right.\right.$$

$$\left.\left. + w_r(x) \left[\frac{\partial f_r(x)}{\partial r} - \mathrm{div}\, b_r^\delta(x) f_r(x) \right] \right) dx \right] dr$$

$$= \int_s^t \mathbb{E}\left[\int_{\mathbb{R}^d} w_r(x) \left(\frac{\partial f_r(x)}{\partial r} - \mathrm{div}\,\mathrm{div}(\widetilde{a}_r(x) f_r(x)) - b_r^\delta(x) \cdot Df_r(x) \right.\right.$$

$$- \operatorname{div} b_r^\delta(x) f_r(x) \bigg) dx - \Big\langle \zeta_r^\delta, f_r \Big\rangle \bigg] dr$$

$$= \int_s^t \mathbb{E} \bigg[\int_{\mathbb{R}^d} w_r(x) \bigg(\frac{\partial f_r(x)}{\partial r} - \operatorname{tr}[\widetilde{a}_r(x) D^2 f_r(x)] - (b_r^\delta(x) + 2 \operatorname{div} \widetilde{a}_r(x)) \cdot Df_r(x)$$

$$- (\operatorname{div} b_r^\delta(x) + \operatorname{div} \operatorname{div} \widetilde{a}_r(x)) f_r(x) \bigg) dx - \Big\langle \zeta_r^\delta, f_r \Big\rangle \bigg] dr.$$

Then, if we choose λ, $\mathcal{K}$, and $C > 0$ sufficiently large, depending only on $\|\widetilde{a}\|_{C^{1,1}}$, the local bounds for $D_p \widetilde{H}^1$ and $D_{pp}^2 \widetilde{H}^1$, and the Lipschitz and semiconcavity bounds for $\widetilde{u}^1$ and $\widetilde{u}^2$, and if we choose $f = \psi_{\lambda,\mathcal{K}}$ from Lemma 3.3 and set

$$\gamma_t := \mathbb{E} \int_{\mathbb{R}^d} w_t(x) \psi_{\lambda,\mathcal{K}}(t,x) dx,$$

we conclude that, for $t \in [0,T]$,

$$\dot{\gamma}_t \geq -C\gamma_t - \Big\langle \zeta_t^\delta, \psi_{\lambda,\mathcal{K},t} \Big\rangle.$$

By the dominated convergence theorem, using the fact that $\psi_{\lambda,\mathcal{K}}(t,\cdot) \in W^{1,1}$, upon sending $\delta \to 0$, we have $\dot{\gamma}_t \geq -C\gamma_t$. Grönwall's inequality gives $\sup_{t \in [0,T]} \gamma_t \leq e^{C(T-t)} \gamma_T$, or, in other words,

$$\sup_{t \in [0,T]} \mathbb{E} \int_{\mathbb{R}^d} \phi(\widetilde{u}_t^1(x) - \widetilde{u}_t^2(x)) \psi_{\lambda,\mathcal{K}}(t,x) dx \leq e^{C(T-t)} \mathbb{E} \int_{\mathbb{R}^d} \phi(\widetilde{G}^1(x) - \widetilde{G}^2(x)) \psi_{\lambda,\mathcal{K}}(t,x) dx.$$

We then take ϕ to be such that $\phi(r) > 0$ if and only if $r < 0$. Using the fact that $\widetilde{G}^1 \geq \widetilde{G}^2$ and $\psi_{\lambda,K} > 0$ in $[0,T] \times \mathbb{R}^d$, we conclude that, with probability one, $\widetilde{u}^1 \geq \widetilde{u}^2$ in $[0,T] \times \mathbb{R}^d$, as desired. $\qquad\square$

3.4. *Stability.* We now discuss stability for (3.1). This result is used later to prove the existence of weak solutions to the mean field games system. It also immediately gives a vanishing viscosity type result relating solutions of (3.1) and solutions of the first order problem considered in [10].

THEOREM 3.3. *Assume $(\widetilde{a}^1, \widetilde{G}^1, \widetilde{H}^1)$ and $(\widetilde{a}^2, \widetilde{G}^2, \widetilde{H}^2)$ satisfy (3.2) and (3.3), and let $(\widetilde{u}^1, M^1)$ and $(\widetilde{u}^2, M^2)$ be the corresponding solutions of (3.1) in the sense of Definition 3.1. Then, for all $R > 0$, there exists $C > 0$ depending on R, d, the Lipschitz and semiconcavity constants for $\widetilde{u}^1$ and $\widetilde{u}^2$, and constants in the given assumptions, such that*

$$\sup_{t \in [0,T]} \mathbb{E} \big\| \widetilde{u}_t^1 - \widetilde{u}_t^2 \big\|_{\infty, B_R}^{d+2}$$

$$\leq C\mathbb{E} \left(\sup_{t \in [0,T]} \big\| \widetilde{a}_t^1 - \widetilde{a}_t^2 \big\|_{C^{0,1}} + \big\| \widetilde{H}^1 - \widetilde{H}^2 \big\|_{\infty, [0,T] \times \mathbb{R}^d \times B_C} + \big\| \widetilde{G}^1 - \widetilde{G}^2 \big\|_\infty \right).$$

PROOF. We argue as in the proof of Theorem 3.2, and let v, M, b, b^δ, ϕ, w, and ζ^δ be defined in the same way. Defining also, for $(t,x) \in [0,T] \times \mathbb{R}^d$,

$$\kappa_t(x) := \widetilde{H}_t^2(x, D\widetilde{u}_t^2(x)) - \widetilde{H}_t^1(x, D\widetilde{u}_t^2(x)),$$

we deduce, by Lemma 3.2, for any $f \in W^{1,1}([0,T] \times \mathbb{R}^d)$ and $0 \le s < t \le T$,

$$
\mathbb{E} \int_{\mathbb{R}^d} w_t(x) f_t(x) dx - \mathbb{E} \int_{\mathbb{R}^d} w_s(x) f_s(x) dx
$$

$$
\ge \int_s^t \mathbb{E} \left[-\left\langle \mathrm{tr}[(\widetilde{a}_r^1 - \widetilde{a}_r^2) D^2 u_r^1] \phi'(v_r) + \mathrm{tr}[\widetilde{a}_r^2 D^2 w_r] + \zeta_r^\delta, f_r \right\rangle \right.
$$

$$
+ \int_{\mathbb{R}^d} \left(\left[-\kappa_r(x) + \mathrm{div}(b_r^\delta(x) w_r(x)) \right] f_r(x) \right.
$$

$$
\left. + w_r(x) \left[\frac{\partial f_r(x)}{\partial r} - \mathrm{div}\, b_r^\delta(x) f_r(x) \right] \right) dx \Bigg] dr
$$

$$
= \int_s^t \mathbb{E} \left[\int_{\mathbb{R}^d} w_r(x) \left(\frac{\partial f_r(x)}{\partial r} - \mathrm{tr}[D^2(\widetilde{a}_r^2(x) f_r(x))] \right. \right.
$$

$$
\left. - b_r^\delta(x) \cdot D f_r(x) - \mathrm{div}\, b_r^\delta(x) f_r(x) \right) dx - \left\langle \mathrm{tr}[(\widetilde{a}_r^1 - \widetilde{a}_r^2) D^2 \widetilde{u}_r^1] \phi'(v_r) + \zeta_r^\delta + \kappa_r, f_r \right\rangle \Bigg] dr
$$

$$
= \int_s^t \mathbb{E} \left[\int_{\mathbb{R}^d} w_r(x) \left(\frac{\partial f_r(x)}{\partial r} - \mathrm{tr}[\widetilde{a}_r^2(x) D^2 f_r(x)] - (b_r^\delta(x) + 2\,\mathrm{div}\,\widetilde{a}_r^2(x)) \cdot D f_r(x) \right. \right.
$$

$$
\left. - (\mathrm{div}\, b_r^\delta(x) + \mathrm{div}\,\mathrm{div}\,\widetilde{a}_r^2(x)) f_r(x) \right) dx
$$

$$
\left. - \left\langle \mathrm{tr}[(\widetilde{a}_r^1 - \widetilde{a}_r^2) D^2 u_r^1] \phi'(v_r) + \zeta_r^\delta + \kappa_r, f_r \right\rangle \right] dr.
$$

We now take $f = \psi_{\lambda, \mathcal{K}}$ as in Lemma 3.2 for $\lambda, \mathcal{K}$ sufficiently large, depending only on the data specified in the statement of the theorem, and set

$$
\gamma_t := \mathbb{E} \int_{\mathbb{R}^d} w_t(x) \psi_{\lambda, \mathcal{K}}(t, x) dx.
$$

Then, upon sending $\delta \to 0$ and using the dominated convergence theorem, we have, for $t \in [0, T]$,

$$
\dot{\gamma}_t \ge -C \gamma_t - \mathbb{E} \left\langle \mathrm{tr}[(\widetilde{a}_t^1 - \widetilde{a}_t^2) D^2 \widetilde{u}_t^1] \phi'(v_t) + \kappa_t, \psi_{\lambda, \mathcal{K}, t} \right\rangle.
$$

We further estimate, for some $C > 0$ as in the statement of the theorem,

$$
|\langle \kappa_t, \psi_{\lambda, \mathcal{K}, t} \rangle| \le C \sup_{(x,p) \in \mathbb{R}^d \times B_C} |\widetilde{H}_t^1(x, p) - \widetilde{H}_t^2(x, p)|,
$$

and

$$
\left\langle \mathrm{tr}[(\widetilde{a}_t^1 - \widetilde{a}_t^2) D^2 \widetilde{u}_t^1] \phi'(v_t), \psi_{\lambda, \mathcal{K}, t} \right\rangle
$$

$$
= -\int_{\mathbb{R}^d} D\widetilde{u}_t^1(x) \cdot \left[\mathrm{div}(\widetilde{a}_t^1(x) - \widetilde{a}_t^2(x)) \phi'(v_t(x)) \psi_{\lambda, \mathcal{K}}(t, x) \right.
$$

$$
+ (\widetilde{a}_t^1(x) - \widetilde{a}_t^2(x)) Dv_t(x) \phi''(v_t(x)) \psi_{\lambda, \mathcal{K}}(t, x)
$$

$$
\left. + (\widetilde{a}_t^1(x) - \widetilde{a}_t^2(x)) D\psi_{\lambda, \mathcal{K}}(t, x) \phi'(v_t(x)) \right] dx.
$$

We then take $\phi(r) = (r_-)^2$. An application of Young's inequality yields, for a different $C > 0$,

$$\dot{\gamma}_t \geq -C\gamma_t - C\mathbb{E}\left(\left\|\widetilde{a}_t^1 - \widetilde{a}_t^2\right\|_{C^{0,1}} + \left\|\widetilde{H}^1 - \widetilde{H}^2\right\|_{\infty,[0,T]\times\mathbb{R}^d\times B_C}\right).$$

Upon applying Grönwall's inequality and repeating the argument after the switching the roles of $\widetilde{u}^1$ and $\widetilde{u}^2$, we obtain

$$\sup_{t\in[0,T]} \mathbb{E}\int_{\mathbb{R}^d} |\widetilde{u}_t^1(x) - \widetilde{u}_t^2(x)|^2 \psi_{\lambda,\mathcal{K}}(t,x)dx$$

$$\leq C\mathbb{E}\left(\left\|\widetilde{G}_1 - \widetilde{G}_2\right\|_\infty^2 + \sup_{t\in[0,T]} \left\|\widetilde{a}_t^1 - \widetilde{a}_t^2\right\|_{C^{0,1}} + \left\|\widetilde{H}^1 - \widetilde{H}^2\right\|_{\infty,[0,T]\times\mathbb{R}^d\times B_C}\right).$$

The desired estimate now follows from Lemma 3.5 with $p = 2$, and the boundedness of $\widetilde{G}^1$ and $\widetilde{G}^2$. $\qquad\square$

3.5. *Optimal control representation.* We finish this section by stating an optimal control representation of the solution. The result, and proof, are similar to the first-order case in [10, Proposition 2.7].

We assume without loss of generality that the probability space $(\Omega, \mathbb{P})$ supports an m-dimensional Brownian motion $(B_t)_{t\in[0,T]}$ independent of $\mathbb{F}$. We then define the set of admissible controls

$$\mathcal{C} := \{\alpha \in L^\infty([0,T]\times\Omega;\mathbb{R}^d) : \alpha \text{ is progressively measurable}$$

$$\text{with respect to the filtration generated by } B\}.$$

Given $\alpha \in \mathcal{C}$, we denote by $X^{\alpha,t,x} : [t,T]\times\Omega^* \to \mathbb{R}^d$ the unique solution of the stochastic differential equation

(3.20)
$$\begin{cases} dX_s^{\alpha,t,x} = \alpha_s ds + \widetilde{\sigma}_s(X_s^{\alpha,t,x})dB_s, & t \leq s \leq T, \\ X_t^{\alpha,t,x} = x. \end{cases}$$

Observe that $X^{\alpha,t,x}$ is adapted to the filtration generated by $\mathbb{F}$ and B.

PROPOSITION 3.1. *The unique solution $\widetilde{u}$ of* (3.1) *has the representation*

$$\widetilde{u}_t(x) = \operatorname*{ess\,inf}_{\alpha\in\mathcal{C}} \mathbb{E}\left[\int_t^T \widetilde{H}_s^*(X_s^{t,x,\alpha},\alpha_s)ds + \widetilde{G}(X_T^{t,x,\alpha}) \,\bigg|\, \mathbb{F}_t\right].$$

PROOF. Fix $N \in \mathbb{N}$ and let $\widetilde{u}^N$ be as in (3.12). Then, for any $n = 1, 2, \ldots, 2^N$ and $t \in [t_{n-1}^N, t_n^N)$, the standard stochastic control formula (see [29]) gives

$$\widetilde{u}_t^N(x) = \operatorname*{ess\,inf}_{\alpha\in\mathcal{C}} \mathbb{E}\left[\int_t^{t_n^N} \widetilde{H}_s^{N,*}(X_s^{t,x,\alpha,N},\alpha_s)ds + \widetilde{u}_{t_n^N,-}^N(X_{t_n^N}^{t,x,\alpha,N})\right],$$

where $X^{t,x,\alpha,N}$ solves

$$\begin{cases} dX_s^{\alpha,t,x,N} = \alpha_s ds + \widetilde{\sigma}_s^N(X_s^{\alpha,t,x,N})dB_s, & t \leq s \leq T, \\ X_t^{\alpha,t,x,N} = x. \end{cases}$$

The $\mathbb{F}_{t_n^N}$-measurability of H^N and σ^N (and therefore $X^{\alpha,t,x,N}$) on $[t, t_n^N)$ yields, in view of the definition of $u_{t_n^N,-}^N$,

$$\widetilde{u}_t^N(x) = \operatorname*{ess\,inf}_{\alpha \in \mathcal{C}} \mathbb{E}\left[\int_t^{t_n^N} \widetilde{H}_s^{N,*}(X_s^{t,x,\alpha,N}, \alpha_s)ds + \widetilde{u}_{t_n^N}^N(X_{t_n^N}^{t,x,\alpha,N}) \,\Big|\, \mathbb{F}_t^N\right],$$

and an inductive argument leads to

$$(3.21) \qquad \widetilde{u}_t^N(x) = \operatorname*{ess\,inf}_{\alpha \in \mathcal{C}} \mathbb{E}\left[\int_t^T \widetilde{H}_s^{N,*}(X_s^{t,x,\alpha,N}, \alpha_s)ds + \widetilde{G}(X_T^{t,x,\alpha,N}) \,\Big|\, \mathbb{F}_t^N\right].$$

There exists $C > 0$ depending only on T and the Lipschitz constant of $\widetilde{\sigma}$ in the $\mathbb{R}^d$-variable such that, uniformly in $(t,x) \in [0,T] \times \mathbb{R}^d$, $\alpha \in \mathcal{C}$, $N \in \mathbb{N}$, and $s \in [t,T]$,

$$\mathbb{E}|X_s^{t,x,\alpha,N} - X_s^{t,x,\alpha}|^2 \le C\mathbb{E}\int_0^T \left\|\widetilde{\sigma}_s^N - \widetilde{\sigma}_s\right\|_\infty^2 ds \xrightarrow{N \to \infty} 0.$$

By Lemma 3.4, $\left\|D\widetilde{u}^N\right\|_{\infty, \Omega \times [0,T] \times \mathbb{R}^d}$ is bounded uniformly in N. It follows that there exists a deterministic $M > 0$ independent of N such that the essential infimum in (3.21) can be restricted to $\alpha \in \mathcal{C}$ satisfying $\|\alpha\|_\infty \le M$. Using the uniform continuity of $\widetilde{\sigma}$ on $[0,T] \times \mathbb{R}^d$ and $\widetilde{H}^*$ on $[0,T] \times \mathbb{R}^d \times B_M$, we may take the limit as $N \to \infty$ in (3.21), which yields the result. $\qquad\square$

4. The MFG system.

We now consider the MFG system (1.1). As discussed in the introduction, this is done by studying the transformed system

$$(4.1) \quad \begin{cases} d\widetilde{u}_t = \left[-\operatorname{tr}(\widetilde{a}_t(x, \widetilde{m}_t)D^2\widetilde{u}_t) + \widetilde{H}_t(x, D\widetilde{u}_t, \widetilde{m}_t)\right]dt + dM_t \text{ in } (0,T) \times \mathbb{R}^d, \\ d\widetilde{m}_t = \operatorname{div}\left[\operatorname{div}(\widetilde{a}_t(x, \widetilde{m}_t)\widetilde{m}_t) + \widetilde{m}_t D_p\widetilde{H}_t(x, D\widetilde{u}_t, \widetilde{m}_t)\right]dt \text{ in } (0,T) \times \mathbb{R}^d, \\ \widetilde{u}_T(x) = \widetilde{G}(x, \widetilde{m}_T), \quad \widetilde{m}_0 = \bar{m}_0 \text{ in } \mathbb{R}^d. \end{cases}$$

We begin by proving a general existence result, Theorem 4.1, under the assumption that $\widetilde{a}, \widetilde{G}$, and $\widetilde{H}$ satisfy the same conditions as in Section 3, uniformly in the m-variable and Lipschitz continuous with respect to the Wasserstein distance. Uniqueness of probabilistically strong solutions is then established under an additional monotonicity condition in Theorem 4.2.

We remind the reader that (4.1) is formally derived from the original system (1.1) through the relations given, for $(t,x,p,m) \in [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$, by

$$(4.2) \qquad \widetilde{u}_t(x) = u_t(x + \sqrt{2\beta}W_t), \qquad \widetilde{m}_t = (\operatorname{Id} - \sqrt{2\beta}W_t)_\sharp m_t,$$

$$(4.3) \qquad \widetilde{a}_t(x,m) = a_t(x + \sqrt{2\beta}W_t, (\operatorname{Id} + \sqrt{2\beta}W_t)_\sharp m),$$

$$(4.4) \qquad \widetilde{H}_t(x,p,m) = H_t(x + \sqrt{2\beta}W_t, p, (\operatorname{Id} + \sqrt{2\beta}W_t)_\sharp m),$$

$$(4.5) \qquad \widetilde{G}(x,m) = G(x + \sqrt{2\beta}W_T, (\operatorname{Id} + \sqrt{2\beta}W_T)_\sharp m),$$

and

$$(4.6) \quad \widetilde{v}_t(x) = v_t(x - \sqrt{2\beta}W_t) - \sqrt{2\beta}Du_t(x - \sqrt{2\beta}W_t), \quad \text{and} \quad M_t = \int_0^t \widetilde{v}_s dW_s.$$

Motivated by the theory of weak and strong solutions to stochastic ordinary differential equations, we define a notion of probabilistically weak solutions of (4.1). More precisely, a weak solution is characterized not only by the unknowns $\widetilde{u}$, M, and $\widetilde{m}$, but also the filtered probability space and Wiener process W.

DEFINITION 4.1. We say that $(\Omega, \mathbb{F}, \mathbb{P}, W, \widetilde{u}, M, \widetilde{m})$ is a weak solution of (4.1) if

(a) $(\Omega, \mathbb{F}, \mathbb{P})$ is a filtered probability space, with a complete right-continuous filtration $\mathbb{F}$,

(b) The random variable $\widetilde{m}$ takes values in $C([0, T], \mathcal{P}_2(\mathbb{R}^d))$ with $\mathbb{P}$-probability one, $\widetilde{m}$ is adapted to the filtration $\mathbb{F}$, and W is a d-dimensional $\mathbb{F}$-Brownian motion,

(c) $(\widetilde{u}, M)$ satisfies, in the sense of Definition 3.1, the Hamilton-Jacobi equation

$$d\widetilde{u}_t = \big[-\operatorname{tr}(\widetilde{a}_t(x, \widetilde{m}_t) D^2 \widetilde{u}_t) + \widetilde{H}_t(x, D\widetilde{u}_t, \widetilde{m}_t) \big] dt + dM_t \ \ \text{in} \ \ (0, T) \times \mathbb{R}^d,$$

(d) $\mathbb{P}$-a.s. and in the sense of distributions, $\widetilde{m}$ solves the Fokker-Planck equation

$$d\widetilde{m}_t = \operatorname{div}\big[\operatorname{div}(\widetilde{a}_t(x, \widetilde{m}_t) \widetilde{m}_t) + \widetilde{m}_t D_p \widetilde{H}_t(x, D\widetilde{u}_t, \widetilde{m}_t) \big] dt \ \ \text{in} \ \ (0, T) \times \mathbb{R}^d,$$

and

(e) $\widetilde{m}_0 = \bar{m}_0$ and $\widetilde{u}_T = \widetilde{G}(\cdot, \widetilde{m}_T)$ $\mathbb{P}$-a.s.

We say that the solution is strong if $(\widetilde{u}, M, \widetilde{m})$ is adapted to the complete filtration generated by W.

REMARK 4.1. In this paper, we give meaning to the original system (1.1) by solving (4.1). We do not explore whether the weak solution of the transformed system leads to some notion of solution (u, v, m), and indeed, it is not clear whether a process $\widetilde{v}$ as in (4.6) can be extracted from a probabilistically weak solution $(\widetilde{u}, M, \widetilde{m})$. On the other hand, in the theory of mean field games, one is typically interested in the value function u and the density of agents m, which can be recovered as follows:

$$u_t(x) = \widetilde{u}_t(x + \sqrt{2\beta} W_t), \quad m_t = (\operatorname{Id} - \sqrt{2\beta} W_t)_{\#} \widetilde{m}_t.$$

REMARK 4.2. In view of (4.6) and Definition 4.1, the term $\operatorname{div}_x v$ appearing in (1.1) would be a distribution which is the derivative of a measure in $\mathcal{M}_{\mathrm{loc}}$. The study of the transformed system (4.1) then has another advantage: not only is the forward Fokker-Planck equation nonstochastic, but also, all terms in the Hamilton-Jacobi-Bellman are, at worst, distributions belonging to $\mathcal{M}_{\mathrm{loc}} \cap W^{-1, \infty}$.

REMARK 4.3. It does not appear to be the case that the transformations (4.2) - (4.6) have a regularizing or averaging effect, a phenomenon which has been exploited for ODEs with rough or stochastic forcing [14, 19, 13]. This is because u and m are themselves stochastic processes adapted to the same filtration as W.

4.1. *Existence of a weak solution.* We are able to prove the existence of a weak solution, that is, the tuple $(\Omega, \mathbb{F}, \mathbb{P}, W, \widetilde{u}, M, \widetilde{m})$ satisfying Definition 4.1, under general assumptions on the deterministic functions a, H, and G, recalling that $\widetilde{a}$, $\widetilde{H}$, and $\widetilde{G}$ are then defined with the constructed Wiener process W through (4.3), (4.4), and (4.5).

In the case where $a \equiv 0$, treated in [10], or in the nondegenerate setting of [8], existence and uniqueness of a strong solution is established by assuming monotonicity and a separated structure from the beginning. Our existence result is more general than these even when a is either zero or nondegenerate, since we generalize to nonseparated dependence, and assume nothing more about the dependence on the measure variable m except mild regularity.

More precisely, we will assume that

(4.7)
$$
\begin{cases}
a.(\cdot, m) : [0,T] \times \mathbb{R}^d \to \mathbb{S}^d,\ G(\cdot, m) : \mathbb{R}^d \to \mathbb{R}, \\[2pt]
\qquad \text{and } H.(\cdot, \cdot, m) : [0,T] \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R} \\[2pt]
\text{satisfy (1.9), (1.10), (1.11), and (1.12) uniformly over } m \in \mathcal{P}_2(\mathbb{R}^d), \\[2pt]
\text{and, for all } R > 0,\ \text{uniformly over } (t,x,p) \in [0,T] \times \mathbb{R}^d \times B_R, \\[2pt]
m \mapsto (\sigma_t(x,m), G(x,m), H_t(x,p,m)) \text{ is Lipschitz in the } \mathcal{P}_2(\mathbb{R}^d) \text{ distance.}
\end{cases}
$$

In order to exploit compactness in $\mathcal{P}_2(\mathbb{R}^d)$, we also assume that the initial distribution satisfies

(4.8)
$$
\bar{m}_0 \in \mathcal{P}_p(\mathbb{R}^d) \text{ for some } p > 2 \text{ and has a bounded density.}
$$

THEOREM 4.1.   *Assume (4.7) and (4.8). Then there exists a weak solution of (4.1).*

4.1.1. *The discretized system.*   As a first step, we construct a (strong) solution of a discretized version of (4.1). More precisely, rather than discretizing the time and state spaces, we directly discretize the canonical probability space of paths. This leads to a fixed point problem on a metric space which, due to the finiteness of the probability space, is now compact, for which Kakutani's fixed point theorem is well-suited.

Let $\Omega^0 = C([0,T], \mathbb{R}^d)$, $\mathbb{F}^0 = (\mathbb{F}^0_t)_{t \in [0,T]}$ be the filtration generated by the canonical process $W \in \Omega^0$, and $\mathbb{P}^0$ be the Wiener measure on Ω^0, so that W is a Brownian motion under $\mathbb{P}^0$.

The discretization of Ω^0 is done through a certain projection map onto piecewise affine paths, whose slopes lie in a finite but possibly large set. The properties of this construction are outlined in the next lemma, whose proof we omit. A similar procedure is discussed in detail in the work of Carmona, Delarue, and Lacker [12, 2.4.2].

LEMMA 4.1.   *For $n \in \mathbb{N}$, there exists a map $\pi^n : \Omega^0 \to \Omega^0$ such that the following hold:*

(i) *π^n is nonanticipative and Borel measurable,*
(ii) *for any $x \in \Omega^0$ and $i = 0, 1, 2, \ldots, 2^n - 1$, $\pi^n(x)$ is affine on*

$$
[t^n_i, t^n_{i+1}] := \left[\frac{Ti}{2^n}, \frac{T(i+1)}{2^n} \right],
$$

(iii) *the increments $(\pi^n(x)(t^n_{i+1}) - \pi^n(x)(t^n_i))_{i=0,1,2,\ldots,2^n-1}$ belong to a finite set depending only on n, independent of x, and,*
(iv) *if $\mathbb{F}^n = (\mathbb{F}^n_t)_{t \in [0,T]}$ denotes the piecewise-constant, right-continuous filtration generated by $\pi^n(W)$, then*

(4.9)
$$
\lim_{n \to \infty} \mathbb{E}_{\mathbb{P}^0} \left[\sup_{t \in [0,T]} |\pi^n(W)_t - W_t| \right] = 0 \quad \text{and}
$$

$$
\bigcup_{n \in \mathbb{N}} \mathbb{F}^n_t = \mathbb{F}^0_t \quad \text{for all } t \in [0,T].
$$

We now construct the discretized variant of (4.1) by defining $\widetilde{a}^n$, $\widetilde{H}^n$, and $\widetilde{G}^n$ from a, H and G just as in (4.3), (4.4) and (4.5), replacing W in each with the discrete projection

$\pi^n(W)$. We then seek a solution $(\widetilde{u}^n, M^n, \widetilde{m}^n)$ of

(4.10)
$$\begin{cases} d\widetilde{u}_t^n = \left[-\operatorname{tr}(\widetilde{a}_t^n(x, m_t^n) D^2 \widetilde{u}_t^n) + \widetilde{H}_t^n(x, D\widetilde{u}_t^n, \widetilde{m}_t^n) \right] dt + dM_t^n \ \text{ in } \ (0,T) \times \mathbb{R}^d, \\ d\widetilde{m}_t^n = \operatorname{div} \left[\operatorname{div}(\widetilde{a}_t^n(x, \widetilde{m}_t^n) \widetilde{m}_t^n) + \widetilde{m}_t^n D_p \widetilde{H}_t^n(x, D\widetilde{u}_t^n, \widetilde{m}_t^n) \right] dt \ \text{ in } \ (0,T) \times \mathbb{R}^d, \\ \widetilde{m}_0^n = \bar{m}_0, \ \ \widetilde{u}_T^n(x) = \widetilde{G}^n(x, \widetilde{m}_T^n) \ \text{ in } \ \mathbb{R}^d. \end{cases}$$

PROPOSITION 4.1.   *There exists a solution $(\widetilde{u}^n, M^n, \widetilde{m}^n)$ of* (4.10) *adapted to the discrete filtration* $\mathbb{F}^n$.

Proposition 4.1 is proved by setting up an appropriate fixed point problem. We define, for $R > 0$ (recall that $p > 2$ is as in (4.8)),

$$\mathcal{M}_R^n := \Bigg\{ m \in L^\infty\big(\Omega^0; C([0,T], \mathcal{P}_2(\mathbb{R}^d)) \cap L^\infty([0,T] \times \mathbb{R}^d)\big), \ m \text{ is } \mathbb{F}^n\text{-adapted}, \ m_0 = \bar{m}_0,$$

$$\text{and } \|m\|_{\infty, \Omega^0 \times [0,T] \times \mathbb{R}^d} + \operatorname*{ess\,sup}_{\omega \in \Omega^0} \sup_{t \in [0,T]} \int_{\mathbb{R}^d} |x|^p m_t(x) dx$$

$$+ \operatorname*{ess\,sup}_{\omega \in \Omega^0} \sup_{s,t \in [0,T], s \neq t} \frac{\mathbf{d}_2(m_s, m_t)}{|s-t|^{1/2}} \le R \Bigg\},$$

along with the metric

$$\mathbf{d}_{\mathcal{M}_R^n}(m, m') = \mathbb{E}\left[\sup_{t \in [0,T]} \mathbf{d}_2(m_t, m_t') \right].$$

The space $\mathcal{M}_R^n$ contains the constant process $\bar{m}_0$, and is thus nonempty, as long as

$$R \ge \|\bar{m}_0\|_\infty + \int_{\mathbb{R}^d} |x|^p \bar{m}_0(x) dx.$$

Moreover, $(\mathcal{M}_R^n, \mathbf{d}_{\mathcal{M}^n})$ is a convex and compact metric space. Indeed, because the processes are $\mathbb{F}^n$-adapted, they may be identified with random variables defined over a finite probability space, and they take values in a compact subset of $C([0,T], \mathcal{P}_2(\mathbb{R}^d))$.

For a fixed $\widetilde{m} \in \mathcal{M}_R^n$, we consider the backward stochastic Hamilton-Jacobi equation

(4.11)
$$\begin{cases} d\widetilde{u}_t^{(\widetilde{m})} = \left[-\operatorname{tr}(\widetilde{a}_t^n(x, \widetilde{m}_t) D^2 \widetilde{u}_t^{(\widetilde{m})}) + \widetilde{H}_t^n(x, D\widetilde{u}_t^{(\widetilde{m})}, \widetilde{m}_t) \right] dt \\ \qquad\qquad + dM_t^{(\widetilde{m})} \ \text{ in } \ (0,T) \times \mathbb{R}^d, \\ \widetilde{u}_T^{(\widetilde{m})}(x) = \widetilde{G}^n(x, \widetilde{m}_T) \ \text{ in } \ \mathbb{R}^d. \end{cases}$$

LEMMA 4.2.   *For any $\widetilde{m} \in \mathcal{M}_R^n$, there exists a unique solution $(\widetilde{u}^{(\widetilde{m})}, M^{(\widetilde{m})})$ of* (4.11)*, and there exists a constant $C > 0$ that is independent of R and $\widetilde{m} \in \mathcal{M}_R^n$ such that, with probability one,*

$$\left\| \widetilde{u}^{(\widetilde{m})} \right\|_\infty + \left\| D\widetilde{u}^{(\widetilde{m})} \right\|_\infty + \operatorname{ess\,sup} m_+(D^2 \widetilde{u}^{(\widetilde{m})}) \le C.$$

Moreover, the map

$$(\mathcal{M}_R^n, \mathbf{d}_{\mathcal{M}^n}) \ni \widetilde{m} \mapsto D\widetilde{u}^{(\widetilde{m})} \in L^1(\Omega^0, L^1_{\operatorname{loc}}([0,T] \times \mathbb{R}^d))$$

is continuous.

PROOF. In view of (4.7), the random variables

$$(t, x) \mapsto \left(\widetilde{a}_t^n(x, \widetilde{m}_t), \widetilde{H}_t^n(x, \cdot, \widetilde{m}_t), \widetilde{G}_t^n(x, \widetilde{m}_t) \right)$$

satisfy (3.3) for the filtered probability space $(\Omega^0, \mathbb{F}^n, \mathbb{P}^0)$, with constants that are independent of R and $\widetilde{m}$. The existence and uniqueness of a solution satisfying the desired bounds is then a consequence of Theorems 3.1 and 3.2, and the continuity statement follows from the uniform Lipschitz continuity of $(\widetilde{a}^n, \widetilde{H}^n, \widetilde{G}^n)$ in the $\mathcal{P}_2$-variable, Theorem 3.3, Lemma 2.1(d), and the bounded convergence theorem. □

LEMMA 4.3. *Fix* $\widetilde{m} \in \mathcal{M}_R^n$, *and let* $(\widetilde{u}^{(\widetilde{m})}, M^{(\widetilde{m})})$ *be as in Lemma 4.2. Then there exists a constant* $\widetilde{R} > 0$, *independent of* R *and* $\widetilde{m}$, *and an* $\mathbb{F}^n$-*adapted process* $\mu \in C([0, T], \mathcal{P}_2(\mathbb{R}^d)) \cap L^\infty([0, T] \times \mathbb{R}^d)$ *such that, with probability one,*

$$\|\mu\|_{\infty, [0,T] \times \mathbb{R}^d} + \sup_{t \in [0,T]} \int_{\mathbb{R}^d} |x|^p \mu_t(x) dx + \sup_{s,t \in [0,T], s \neq t} \frac{\mathbf{d}_2(\mu_s, \mu_t)}{|s-t|^{1/2}} \leq \widetilde{R},$$

and, in the sense of distributions,

$$(4.12) \quad \begin{cases} d\mu_t = \mathrm{div} \left[\mathrm{div}(\widetilde{a}_t^n(x, \widetilde{m}_t)\mu_t) + \mu_t D_p \widetilde{H}_t^n(x, D\widetilde{u}_t^{(\widetilde{m})}, \widetilde{m}_t) \right] dt \ \ in \ \ (0, T) \times \mathbb{R}^d, \\ \mu_0 = \bar{m}_0 \ \ in \ \mathbb{R}^d. \end{cases}$$

PROOF. For $\varepsilon > 0$, let ρ_ε be a standard mollifier as in (1.15), and set

$$\widetilde{u}^{(\widetilde{m}),\varepsilon} = \rho_\varepsilon * \widetilde{u}^{(\widetilde{m})} \quad \text{and} \quad b_t^\varepsilon(x) := -D_p \widetilde{H}_t^n(x, D\widetilde{u}_t^{(\widetilde{m}),\varepsilon}(x), \widetilde{m}_t).$$

We then consider the regularized equation

$$(4.13) \quad \begin{cases} \partial_t \mu_t^\varepsilon = \mathrm{div} \left[\varepsilon D\mu_t^\varepsilon + \mathrm{div}(\widetilde{a}_t^n(x, \widetilde{m}_t)\mu_t^\varepsilon) - \mu_t^\varepsilon b_t^\varepsilon \right] \ \ in \ \ (0, T) \times \mathbb{R}^d, \\ \mu_0^\varepsilon = \bar{m}_0 \ \ in \ \mathbb{R}^d. \end{cases}$$

With probability one, (4.13) is a uniformly elliptic equation with coefficients which are smooth and bounded in space and measurable in time. Therefore, there exists a unique solution in the sense of distributions, which is adapted to $\mathbb{F}^n$, since $\widetilde{a}^n$, $\widetilde{H}^n$ and $\widetilde{u}^{(\widetilde{m})}$ are. Standard arguments yield $\widetilde{R}$ independent of R, $\widetilde{m}$, and ε such that

$$\sup_{t \in [0,T]} \int_{\mathbb{R}^d} |x|^p \mu_t^\varepsilon(x) dx + \sup_{s,t \in [0,T], s \neq t} \frac{\mathbf{d}_2(\mu_s^\varepsilon, \mu_t^\varepsilon)}{|s-t|^{1/2}} \leq \widetilde{R}.$$

By Lemma 4.2, $\widetilde{u}^{\widetilde{m},\varepsilon}$ is uniformly Lipschitz continuous and semiconcave in space. This, along with the uniform convexity of $\widetilde{H}^n$ in the gradient variable, implies that there exists $C > 0$ independent of R, $\widetilde{m}$, and ε such that $\mathrm{div}_x b^\varepsilon \geq -C$. The maximum principle then implies that μ^ε is bounded in $L^\infty(\Omega^0 \times [0, T] \times \mathbb{R}^d)$ independent of ε, R, and $\widetilde{m}$, and so, enlarging $\widetilde{R}$ if necessary,

$$\|\mu^\varepsilon\|_{\infty, [0,T] \times \mathbb{R}^d} + \sup_{t \in [0,T]} \int_{\mathbb{R}^d} |x|^p \mu_t^\varepsilon(x) dx + \sup_{s,t \in [0,T], s \neq t} \frac{\mathbf{d}_2(\mu_s^\varepsilon, \mu_t^\varepsilon)}{|s-t|^{1/2}} \leq \widetilde{R}.$$

Exploiting the finiteness of the probability space $(\Omega^0, \mathbb{F}^n, \mathbb{P}^0)$, we then extract a subsequence of (μ^ε) that converges weakly-* in $L^\infty(\Omega^0 \times [0, T] \times \mathbb{R}^d)$ to some μ, which, with probability one, is a solution of (4.12) with the desired properties. □

We now combine Lemmas 4.2 and 4.3 to set up the fixed point argument.

PROOF OF PROPOSITION 4.1. Set $R = \widetilde{R}$, where $\widetilde{R}$ is as in Lemma 4.3 (this is allowed because $\widetilde{R}$ is independent of R). Given $\widetilde{m} \in \mathcal{M}_R^n$, we let $\mathcal{E}(\widetilde{m})$ be the set of processes μ that solve (4.12) and satisfy the properties laid out in Lemma 4.3. This defines a map taking values in convex subsets of $\mathcal{M}_R^n$:

$$\mathcal{M}_R^n \ni \widetilde{m} \mapsto \mathcal{E}(\widetilde{m}) \in 2^{\mathcal{M}_R^n},$$

and, in view of Lemma 4.2 and the uniform Lipschitz continuity of the data in the $\mathcal{P}_2$-variable, the graph of $\mathcal{E}$ is closed. By the Kakutani fixed point theorem, there exists $\widetilde{m}^n \in \mathcal{M}_R^n$ such that $\widetilde{m}^n \in \mathcal{E}(\widetilde{m}^n)$. Setting $\widetilde{u}^n = \widetilde{u}^{(\widetilde{m}^n)}$ and $M^n = M^{(\widetilde{m}^n)}$ gives the desired solution. $\qquad\square$

4.1.2. *The passage to the limit.* We begin by embedding the discrete problem above into a larger probability space. For $R = \widetilde{R}$ as in the proof of Proposition 4.1, we define

$$\Omega^* := \Bigg\{ \omega = (m, w) \in C([0, T], \mathcal{P}_2(\mathbb{R}^d) \times \mathbb{R}^d)$$

(4.14) $\qquad \cap\, L^\infty([0, T] \times \mathbb{R}^d, \mathbb{R} \times \mathbb{R}^d) : w_0 = 0,\, m_0 = \bar{m}_0, \quad \text{and}$

$$\|m\|_{\infty, [0,T] \times \mathbb{R}^d} + \sup_{t \in [0,T]} \int_{\mathbb{R}^d} |x|^p m_t(x) dx + \sup_{s, t \in [0,T], s \le t} \frac{\mathbf{d}_2(m_s, m_t)}{|s - t|^{1/2}} \le R \Bigg\}.$$

The set Ω^* is a Polish space with the metric defined, for $\omega = (m, w)$ and $\omega' = (m', w')$, by

$$\mathbf{d}_{\Omega^*}(\omega, \omega') = \sup_{t \in [0,T]} \left(\mathbf{d}_2(m_t, m_t') + |w_t - w_t'| \right).$$

We denote by $\mathbb{F} = (\mathbb{F}_t)_{t \in [0,T]}$ the filtration on Ω^* generated by the canonical process $\omega = (m, w)$.

Let $\mathcal{P}(\Omega^*)$ be the set of Borel probability measures on Ω^* with finite first order moment, endowed with the distance defined, for $P, P' \in \mathcal{P}(\Omega^*)$,

$$\mathbf{d}_{\mathcal{P}(\Omega^*)}(P, P') = \inf_\pi \mathbb{E}_\pi \left[\mathbf{d}_{\bar{\Omega}}(\omega, \omega) \right]$$

where the infimum is taken over the Borel probability measures π on $\Omega^* \times \Omega^*$ with first marginal P and second marginal P'.

By Proposition 4.1, for all $n \in \mathbb{N}$, there exists a solution $(\widetilde{u}^n, M^n, \widetilde{m}^n)$ of the discrete MFG system (4.10), which defines a probability measure P^n on Ω^*:

(4.15) $$P^n := (\widetilde{m}^n, W)_\sharp \mathbb{P}^0 \in \mathcal{P}(\bar{\Omega}).$$

We can likewise express the various data as random variables on Ω^*. Abusing notation slightly, we define, for $n \in \mathbb{N}$, $(t, x, p) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d$, and $\omega = (m, w) \in \Omega^*$,

(4.16) $$\widetilde{a}_t^n(x, \omega) := a_t(x + \sqrt{2\beta} \pi^n(w)_t, (\mathrm{Id} + \sqrt{2\beta} \pi^n(w)_t)_\sharp m_t),$$

(4.17) $$\widetilde{H}_t^n(x, p, \omega) := \widetilde{H}_t(x + \sqrt{2\beta} \pi^n(w)_t, p, (\mathrm{Id} + \sqrt{2\beta} \pi^n(w)_t)_\sharp m_t),$$

(4.18) $$\widetilde{G}^n(x, \omega) = G(x + \sqrt{2\beta} \pi^n(w)_T, (\mathrm{Id} + \sqrt{2\beta} \pi^n(w)_T)_\sharp m_T),$$

(4.19) $$\widetilde{u}_t^n(x, \omega) = \widetilde{u}_t^n(x, \pi^n(w)),$$

and

(4.20) $$M_t^n(\omega) = M_t^n(\pi^n(w)).$$

The filtration $\mathbb{F}^n$ on Ω^0 can be seen as a filtration on Ω^* generated by $(m, w) \mapsto \pi^n(w)$. With this notation, $(\widetilde{u}^n, M^n)$, defined on the filtered probability space $(\Omega^*, \mathbb{F}^n, P^n)$, solves, in the sense of Definition 3.1,

$$
(4.21) \quad
\begin{cases}
d\widetilde{u}_t^n = \left[-\operatorname{tr}(\widetilde{a}_t^n(x, \omega) D^2 \widetilde{u}_t^n) + \widetilde{H}_t^n(x, D\widetilde{u}_t^n, \omega) \right] dt + dM_t^n & \text{in } (0, T) \times \mathbb{R}^d, \\
\widetilde{u}_T^n = \widetilde{G}^n(\cdot, \omega) & \text{in } \mathbb{R}^d.
\end{cases}
$$

The weak solution of (4.1) is obtained by taking a weak limit, as $n \to \infty$, of P^n.

LEMMA 4.4.   *Up to a subsequence, as $n \to \infty$, P^n converges to some $P \in \mathcal{P}(\Omega^*)$ with respect to the distance $\mathbf{d}_{\mathcal{P}(\Omega^*)}$. Moreover, if $\mathbb{F}^P$ denotes the natural right-continuous augmentation of the filtration $\mathbb{F}$ by P-null sets, then the second component w of the canonical process $(m, w) \in \Omega^*$ is a $\mathbb{F}^P$-Brownian motion under P.*

PROOF.   The first marginal of P^n is supported on a compact set of $C([0, T], \mathcal{P}_2(\mathbb{R}^d))$ that is independent of n, and the second marginal is the Wiener measure on $C([0, T], \mathbb{R}^d)$. It follows that $(P^n)_{n \in \mathbb{N}}$ is tight, and so the convergence along a subsequence is then a standard consequence of the fact that weak convergence implies convergence in the $\mathbf{d}_{\mathcal{P}(\Omega^*)}$ distance.

For all n, the random variable $(m, w) \mapsto (w_t)_{t \in [0,T]}$ is a Brownian motion under P^n, and therefore the same is true for P. To finish the proof, it suffices to show that, for all $0 \le s < t \le T$, $w_t - w_s$ is independent of $\mathbb{F}_s^P$.

Fix $N \in \mathbb{N}$, Borel functions $\phi : \mathbb{R}^d \to \mathbb{R}$ and $\psi : (\mathbb{R}^{1+d})^N \to \mathbb{R}$, and $0 \le s_1 < s_2 < \cdots < s_N \le s < t$. For $n \in \mathbb{N}$, we have

$$
(4.22) \quad
\begin{aligned}
& \int_{\Omega^*} \phi(w_t - w_s) \psi((m, w)_{s_1}, (m, w)_{s_2}, \ldots, (m, w)_{s_N}) dP^n(m, w) \\
& = \int_{\Omega^0} \phi(w_t - w_s) \psi\left(\widetilde{m}_{s_1}^n(\cdot, w), w_{s_1}, \widetilde{m}_{s_2}^n(\cdot, w), w_{s_2}, \ldots, \widetilde{m}_{s_N}^n(\cdot, w), w_{s_N} \right) d\mathbb{P}^0(w).
\end{aligned}
$$

For $i = 1, 2, \ldots, N$, $\widetilde{m}_{s_i}^n$ is $\mathbb{F}_{s_i}^n$-measurable, and therefore $\mathbb{F}_{k2^{-n}}^0$-measurable, where k is the largest integer smaller than $2^n s_i$. It follows that $\widetilde{m}_{s_i}^n$ is independent of $w_t - w_s$, and so

$$
\begin{aligned}
& \int_{\Omega^0} \phi(w_t - w_s) \psi\left(\widetilde{m}_{s_1}^n(\cdot, w), w_{s_1}, \widetilde{m}_{s_2}^n(\cdot, w), w_{s_2}, \ldots, \widetilde{m}_{s_N}^n(\cdot, w), w_{s_N} \right) d\mathbb{P}^0(w) \\
& = \int_{\Omega^0} \phi(w_t - w_s) d\mathbb{P}^0(w) \\
& \quad \cdot \int_{\Omega^0} \psi\left(\widetilde{m}_{s_1}^n(\cdot, w'), w'_{s_1}, \widetilde{m}_{s_2}^n(\cdot, w'), w'_{s_2}, \ldots, \widetilde{m}_{s_N}^n(\cdot, w'), w'_{s_N} \right) d\mathbb{P}^0(w') \\
& = \int_{\Omega^*} \phi(w_t - w_s) dP^n(m, w) \int_{\Omega^*} \psi\left((m', w')_{s_1}, (m', w')_{s_2}, \ldots, (m', w')_{s_N} \right) dP^n(m', w').
\end{aligned}
$$

Taking the limit along an appropriate subsequence as $n \to \infty$, (4.22) becomes

$$
\begin{aligned}
& \int_{\Omega^*} \phi(w_t - w_s) \psi((m, w)_{s_1}, (m, w)_{s_2}, \ldots, (m, w)_{s_N}) dP(m, w) \\
& = \int_{\Omega^*} \phi(w_t - w_s) dP(m, w) \int_{\Omega^*} \psi\left((m', w')_{s_1}, (m', w')_{s_2}, \ldots, (m', w')_{s_N} \right) dP(m', w').
\end{aligned}
$$

Because N, $(s_1, s_2, \ldots, s_N)$, ϕ, and ψ were arbitrary, this implies that, under P, $w_t - w_s$ is independent of $\mathbb{F}_s$, and therefore the same is true for $\mathbb{F}_s^P$. $\square$

The rest of this subsection is devoted to proving that the tuple $(\Omega^*, \mathbb{F}^P, P, w, \widetilde{u}, M, m)$ is a weak solution of (4.1), where w and m stand for the processes $(m, w) \mapsto w$ and $(m, w) \mapsto m$. Accordingly, we now set, for $(t, x, p) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d$ and $\omega = (m, w) \in \Omega^*$,

$$(4.23) \qquad \widetilde{a}_t(x, \omega) = a_t(x + \sqrt{2\beta}w_t, (\mathrm{Id} + \sqrt{2\beta}w_t)_\sharp m_t),$$

$$(4.24) \qquad \widetilde{H}_t(x, p, \omega) = H_t(x + \sqrt{2\beta}w_t, p, (\mathrm{Id} + \sqrt{2\beta}w_t)_\sharp m_t),$$

$$(4.25) \qquad \widetilde{G}(x, \omega) = G(x + \sqrt{2\beta}w_T, (\mathrm{Id} + \sqrt{2\beta}w_T)_\sharp m_T).$$

Observe that, in view of (4.7), for all $L > 0$, there exists $C = C_L > 0$ such that, for all $n \in \mathbb{N}$, $\omega, \omega' \in \Omega^*$, and $t \in [0, T]$,

$$
\left\| \widetilde{a}_t^n(\cdot, \omega) - \widetilde{a}_t(\cdot, \omega') \right\|_{C^{0,1}(\mathbb{R}^d)} + \left\| \widetilde{H}_t^n(\cdot, \cdot, \omega) - \widetilde{H}_t(\cdot, \cdot, \omega') \right\|_{\infty, \mathbb{R}^d \times B_L}
$$

$$(4.26) \qquad + \left\| D_p \widetilde{H}_t^n(\cdot, \cdot, \omega) - D_p \widetilde{H}_t(\cdot, \cdot, \omega') \right\|_{\infty, \mathbb{R}^d \times B_L} + \left\| \widetilde{G}^n(\cdot, \omega) - \widetilde{G}(\cdot, \omega') \right\|_{\infty, \mathbb{R}^d}$$

$$
\leq C \left(\mathbf{d}_{\Omega^*}(\omega, \omega') + \|\pi^n(w) - w\|_{\infty, [0, T]} \right).
$$

Now, for the probability space $(\Omega^*, \mathbb{F}^P, P)$, we consider the solution $(\widetilde{u}, M)$, in the sense of Definition 3.1, of the backward stochastic Hamilton-Jacobi-Bellman equation

$$(4.27) \qquad \begin{cases} d\widetilde{u}_t = \left[-\mathrm{tr}(\widetilde{a}_t(x, \omega)D^2\widetilde{u}_t) + \widetilde{H}_t(x, D\widetilde{u}_t, \omega) \right] dt + dM_t & \text{in } (0, T) \times \mathbb{R}^d, \\ \widetilde{u}_T = \widetilde{G}(\cdot, \omega) & \text{in } \mathbb{R}^d. \end{cases}$$

LEMMA 4.5. *For $n \in \mathbb{N}$, let γ^n be an optimal coupling between P^n and P. Then there exists a subsequence $(n_k)_{k \in \mathbb{N}}$ such that, for all $R > 0$,*

$$(4.28) \qquad \lim_{k \to +\infty} \sup_{t \in [0, T]} \int_{\Omega^* \times \Omega^*} \left[\sup_{x \in B_R} |\hat{u}_t^{n_k}(x, \omega) - \widetilde{u}_t(x, \omega')|^{d+2} \right] d\gamma^{n_k}(\omega, \omega') = 0$$

and

$$(4.29) \qquad \lim_{k \to \infty} \int_{\Omega^* \times \Omega^*} \int_0^T \int_{B_R} |D\widetilde{u}_t^{n_k}(x, \omega) - D\widetilde{u}_t(x, \omega')| dx dt d\gamma^{n_k}(\omega, \omega') = 0.$$

PROOF. For each $n \in \mathbb{N}$, the filtered probability space $(\Omega^* \times \Omega^*, \mathbb{F}^n \otimes \mathbb{F}, \gamma^n)$ and the processes

$$\Omega^* \times \Omega^* \ni (\omega, \omega') \mapsto \left(\widetilde{a}_\cdot^n(\cdot, \omega), \widetilde{H}_\cdot^n(\cdot, \cdot, \omega), \widetilde{G}^n(\cdot, \omega), \widetilde{a}_\cdot(\cdot, \omega'), \widetilde{H}_\cdot(\cdot, \cdot, \omega'), \widetilde{G}(\cdot, \omega') \right)$$

satisfy (3.2) and (3.3), and

$$(\omega, \omega') \mapsto (\widetilde{u}^n(\cdot, \omega), M^n(\cdot, \omega)) \quad \text{and} \quad (\omega, \omega') \mapsto (\widetilde{u}(\cdot, \omega'), M(\cdot, \omega'))$$

solve respectively (4.21) and (4.27) in the sense of Definition 3.1. Theorem 3.3 then gives, for any $R > 0$ and some constants $C, L > 0$ depending on R, d, and the constants the assumptions,

$$\sup_{t \in [0, T]} \int_{\Omega^* \times \Omega^*} \left\| \widetilde{u}_t^n(\cdot, \omega) - \widetilde{u}_t(\cdot, \omega') \right\|_{\infty, B_R}^{d+2} d\gamma^n(\omega, \omega')$$

$$\leq C \int_{\Omega^* \times \Omega^*} \left(\sup_{t \in [0, T]} \left\| \widetilde{a}_t^n(\cdot, \omega) - \widetilde{a}_t(\cdot, \omega') \right\|_{C^{0,1}} + \left\| \widetilde{H}^n(\cdot, \cdot, \omega) - \widetilde{H}(\cdot, \cdot, \omega') \right\|_{\infty, [0, T] \times \mathbb{R}^d \times B_L} \right.$$

$$\left. + \left\| \widetilde{G}^n(\cdot, \omega) - \widetilde{G}(\cdot, \omega') \right\|_\infty \right) d\gamma^n(\omega, \omega').$$

The convergence statement (4.28) now follows from (4.9), (4.26), and Lemma 4.4.

We now define $\boldsymbol{\Omega} := (\Omega^*)^{\mathbb{N}}$, and write $\bar{\omega} = (\omega^0, \omega^1, \omega^2, \ldots) \in \boldsymbol{\Omega}$. We define the cylindrical filtration $\mathbf{F}$ on $\boldsymbol{\Omega}$ by

$$\mathbf{F}_t := \sigma\left((\omega_s^0)_{s \in [0,t]}, (\omega_s^1)_{s \in [0,t]}, \ldots, (\omega_s^n)_{s \in [0,t]}, n = 1, 2, \ldots\right).$$

By the Kolmogorov extension theorem, there exists a unique probability measure $\mathbf{P}$ on $\boldsymbol{\Omega}$ such that, for any $n \in \mathbb{N}$ and Borel measurable $\Phi : (\Omega^*)^{1+n} \to \mathbb{R}$,

$$\int_{\boldsymbol{\Omega}} \Phi(\omega^0, \omega^1, \ldots, \omega^n) d\mathbf{P}(\bar{\omega})$$

$$:= \int_{(\Omega^*)^{1+n}} \Phi(\omega^0, \omega^1, \ldots, \omega^n) d\gamma^1(\omega^1, \omega^0) d\gamma^2(\omega^2, \omega^0) \cdots d\gamma^n(\omega^n, \omega^0).$$

For $(t, x, \bar{\omega}) \in [0, T] \times \mathbb{R}^d \times \boldsymbol{\Omega}$ and $n \in \mathbb{N}$, define

$$\widetilde{U}_t^n(x, \bar{\omega}) := \widetilde{u}_t^n(x, \omega^n) \quad \text{and} \quad \widetilde{U}_t(x, \bar{\omega}) := \widetilde{u}_t(x, \omega^0).$$

Then (4.28) can be rewritten as

$$\lim_{k \to \infty} \sup_{t \in [0,T]} \int_{\boldsymbol{\Omega}} \sup_{x \in B_R} |\widetilde{U}_t^{n_k}(x, \bar{\omega}) - \widetilde{U}_t(x, \bar{\omega})|^{d+2} d\mathbf{P}(\bar{\omega}) = 0,$$

which implies that, for some further subsequence, labeled again as $(n_k)_{k \in \mathbb{N}}$, for all $t \in [0, T]$, $R > 0$, and $\mathbf{P}$-a.s. in $\bar{\omega}$,

$$\lim_{k \to \infty} \sup_{x \in B_R} |\widetilde{U}_t^{n_k}(x, \bar{\omega}) - \widetilde{U}_t(x, \bar{\omega})| = 0.$$

Uniformly over $t \in [0, T]$, $\mathbf{P}$-a.e. $\bar{\omega}$, and $n \in \mathbb{N}$, $\widetilde{U}_t^n(\cdot, \bar{\omega})$ and $\widetilde{U}_t(\cdot, \bar{\omega})$ are globally Lipschitz continuous and semiconcave, and so Lemma 2.1(d) and the dominated convergence theorem imply that, for all $R > 0$,

$$\lim_{k \to \infty} \int_{\boldsymbol{\Omega}} \int_0^T \int_{B_R} |D\widetilde{U}_t^{n_k}(x, \bar{\omega}) - D\widetilde{U}_t(x, \bar{\omega})| dx dt d\mathbf{P}(\bar{\omega}) = 0,$$

which is equivalent to (4.29). $\qquad \square$

PROOF OF THEOREM 4.1. As stated above, we establish the claim of the theorem by proving that $(\Omega^*, \mathbb{F}^P, P, w, \widetilde{u}, M, m)$ is a weak solution of (4.1). Properties (a), (c), and (e) are true by construction, and (b) is a consequence of Lemma 4.4. It remains to show that the process m solves, P-a.s., the Fokker-Planck equation associated to $\widetilde{u}$. To do this, it suffices to show that, for all $\phi \in C_c^1((0, T) \times \mathbb{R}^d)$ and $A \in \mathbb{F}_T^P$,

$$I_A := \int_A \int_0^T \int_{\mathbb{R}^d} m_t(x) \Big[\partial_t \phi_t(x) + \operatorname{tr} \widetilde{a}_t(x, \omega) D^2 \phi_t(x)$$

$$- D\phi_t(x) \cdot D_p \widetilde{H}_t(x, D\widetilde{u}_t(x, \omega), \omega) \Big] dx dt dP(\omega) = 0.$$

The canonical process $(m, w) \mapsto m$ satisfies P^n-a.s. the equation

(4.30) $\qquad \begin{cases} dm_t = \operatorname{div}\big[\operatorname{div}(\widetilde{a}_t^n(x, \omega) m_t) + m_t D_p \widetilde{H}_t^n(x, D\widetilde{u}_t^n, \omega) \big] dt \quad \text{in } (0, T) \times \mathbb{R}^d, \\ m_0 = \bar{m}_0 \quad \text{in } \mathbb{R}^d. \end{cases}$

Let $A' \in \mathbb{F}_T$ and a P-null set $N \subset \Omega$ be such that $A = A' \cup N$. We thus have

$$I_{A'}^n := \int_{A'} \int_0^T \int_{\mathbb{R}^d} m_t(x) \Big[\partial_t \phi_t(x) + \operatorname{tr} \widetilde{a}_t^n(x, \omega) D^2 \phi_t(x)$$

$$- D\phi_t(x) \cdot D_p \widetilde{H}_t^n(x, D\widetilde{u}_t^n(x, \omega), \omega) \Big] dx \, dt \, dP^n(\omega) = 0.$$

Let $R > 0$ be such that $\operatorname{supp} \phi \subset [0, T] \times B_R$. By (4.26), there exists a constant $C > 0$ such that, for all $n \in \mathbb{N}$,

$$|I_{A'}| = |I_{A'} - I_{A'}^n| \leq C \|\phi\|_{C_t^1 C_x^2} \bigg(d_{\mathcal{P}(\Omega^*)}(P^n, P) + \mathbb{E}^0 \|\pi_n(w) - w\|_{\infty, [0, T]}$$

$$+ \int_{\Omega^* \times \Omega^*} \int_0^T \int_{B_R} |D\widetilde{u}_t^n(x, \omega) - D\widetilde{u}_t(x, \omega')| d\gamma^n(\omega, \omega') \bigg),$$

where, as in Lemma 4.5, γ^n is an optimal coupling between P^n and P. Sending $n \to \infty$ along the appropriate subsequence, we conclude from (4.9), Lemma 4.4, and Lemma 4.5 that $I_{A'} = 0$, and therefore $I_A = I_{A' \cup N} = 0$, because $P(N) = 0$. $\qquad \square$

4.2. Uniqueness.

A weak solution in the sense of Definition 4.1 can be identified with the law of $(\widetilde{m}, W)$ on the probability space Ω^* defined in (4.14). Indeed, $(\widetilde{u}, M)$ can then be recovered by solving the backward stochastic HJB equation with the theory from Section 3. This motivates the following definition.

DEFINITION 4.2. Uniqueness in law holds for (4.1) if, for any two weak solutions $(\Omega_i, \mathbb{F}_i, \mathbb{P}_i, W_i, \widetilde{u}_i, M_i, \widetilde{m}_i)_{i=1,2}$, we have

$$(\widetilde{m}_1, W_1)_\sharp \mathbb{P}_1 = (\widetilde{m}_2, W_2)_\sharp \mathbb{P}_2 \quad \text{in } \mathcal{P}(\Omega^*).$$

Pathwise uniqueness holds if, for two solutions $(\widetilde{u}_i, M_i, \widetilde{m}_i)_{i=1,2}$ defined on the same probability space with a given Wiener process W, we have $\widetilde{u}_1 = \widetilde{u}_2$, $M_1 = M_2$, and $\widetilde{m}_1 = \widetilde{m}_2$ with probability one.

Under further structural assumptions, namely, the well-known Lasry-Lions monotonicity conditions [26], we now establish pathwise uniqueness for the system (4.1).

In addition to (4.7) and (4.8), we assume that (abusing notation)

(4.31) $$a_t(x, m) = a_t(x) \text{ is independent of } m \in \mathcal{P}_2(\mathbb{R}^d),$$

(4.32) $$H_t(x, p, m) = H_t(x, p) - F_t(x, m) \quad \text{for } (t, x, p, m) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d),$$

and the coupling functions F and G are strictly monotone, that is,

(4.33) $$\begin{cases} \text{for all } m, m' \in \mathcal{P}_2(\mathbb{R}^d) \text{ and } t \in [0, T], \\[2mm] \displaystyle\int_{\mathbb{R}^d} (F_t(x, m) - F_t(x, m'))(m - m')(dx) \geq 0 \text{ and} \\[2mm] \displaystyle\int_{\mathbb{R}^d} (G(x, m) - G(x, m'))(m - m')(dx) \geq 0, \\[2mm] \text{with equality in either holding if and only if } m = m'. \end{cases}$$

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

$$\widetilde{F}_t(x, m) := F_t(x + \sqrt{2\beta}W_t, (\mathrm{Id} + \sqrt{2\beta}W_t)_\sharp m), \tag{4.34}$$

the MFG system (4.1) then becomes

$$\begin{cases} d\widetilde{u}_t = \left[-\operatorname{tr}(\widetilde{a}_t(x)D^2\widetilde{u}_t) + \widetilde{H}_t(x, D\widetilde{u}_t) - \widetilde{F}_t(x, \widetilde{m}_t) \right] dt + dM_t & \text{in } (0, T) \times \mathbb{R}^d, \\[2mm] d\widetilde{m}_t = \operatorname{div}\left[\operatorname{div}(\widetilde{a}_t(x)\widetilde{m}_t) + \widetilde{m}_t D_p\widetilde{H}_t(x, D\widetilde{u}_t) \right] dt & \text{in } (0, T) \times \mathbb{R}^d, \\[2mm] \widetilde{u}_T(x) = \widetilde{G}(x, \widetilde{m}_T), \quad \widetilde{m}_0 = \bar{m}_0 \ \text{ in } \mathbb{R}^d. \end{cases} \tag{4.35}$$

THEOREM 4.2.   *Assume (4.7), (4.8), and (4.31) - (4.33). Then pathwise uniqueness holds for (4.35) in the sense of Definition 4.2, and, moreover, every weak solution is a strong solution.*

The argument is an adaptation of those in [25, 26], similar to [10]. Informally, one would like to compute

$$d \int_{\mathbb{R}^d} (\widetilde{u}_t^1(x) - \widetilde{u}_t^2(x))(\widetilde{m}_t^1(x) - \widetilde{m}_t^2(x)) dx,$$

take expectations to cancel the martingale terms, and invoke the convexity of H and the monotonicity of F and G to show that $\widetilde{m}^1 = \widetilde{m}^2$. The main subtlety, which does not arise in the first order setting of [10], is the possible discontinuity of $\widetilde{m}^1$ and $\widetilde{m}^2$ in the state variable, which prevents them from being used as test functions in the distributional equality in Definition 3.1. At the same time, $\widetilde{u}^1$ and $\widetilde{u}^2$ fail to be $C^{1,1}$ in the state variable in general, and so cannot be used as test functions for the Fokker-Planck equation.

We overcome this issue by introducing a standard mollifier and cutoff, and the resulting commutators vanish in the limit precisely because $D^2\widetilde{u}^1$ and $D^2\widetilde{u}^2$ are locally finite Radon measures.

PROOF OF THEOREM 4.2.   It is enough to prove the pathwise uniqueness, since then, arguing as in the classic paper of Yamada and Watanabe [34] (see also [12, Proposition 6.1]), invoking the existence of a weak solution by Theorem 4.1, every solution is strong.

Let $(\Omega, \mathbb{F}, \mathbb{P}, W, \widetilde{u}^i, M^i, \widetilde{m}^i)_{i=1,2}$ be two weak solutions, defined on the same probability space with the same Wiener process W. Let $(\rho_\delta)_{\delta>0}$ be as in (1.15) and let $\eta \in C_c^\infty(\mathbb{R}^d)$ be a standard cut-off function equal to 1 in B_1 and 0 outside of B_2, set $\eta_R(x) = \eta(x/R)$ for $x \in \mathbb{R}^d$, $R > 0$, and define

$$\widetilde{m}^{i,\delta} := \widetilde{m}^i * \rho_\delta \quad \text{and} \quad \widetilde{m}^{i,\delta,R} := \widetilde{m}_\delta^i \eta_R, \quad i = 1, 2.$$

Then, for $i = 1, 2$ and $\delta, R > 0$,

$$\partial_t \widetilde{m}_t^{i,\delta,R} = \operatorname{div} \operatorname{div} \left[\widetilde{a}_t(x)\widetilde{m}_t^{i,\delta,R} \right] + \eta_R(x) \operatorname{div} \left[\rho_\delta * \left(D_p\widetilde{H}_t(\cdot, D\widetilde{u}_t^i)\widetilde{m}_t^i \right) \right] \\ + \alpha_t^{i,\delta,R}(x) + \beta_t^{i,\delta,R}(x), \tag{4.36}$$

where

$$\alpha_t^{i,\delta,R}(x) := \eta_R(x) \operatorname{div}_x^2 \int_{\mathbb{R}^d} [\widetilde{a}_t(y) - \widetilde{a}_t(x)]\widetilde{m}_t^i(y)\rho_\delta(x - y) dy$$

and

$$\beta_t^{i,\delta,R}(x) := -\operatorname{div}\left[\widetilde{a}_t(x)\widetilde{m}_t^{i,\delta} \right] \cdot D\eta_R(x) - \widetilde{m}_t^{i,\delta} \operatorname{tr}\left[\widetilde{a}_t(x)D^2\eta_R(x) \right].$$

Set $\alpha^{\delta,R} = \alpha^{1,\delta,R} - \alpha^{2,\delta,R}$ and $\beta^{\delta,R} = \beta^{1,\delta,R} - \beta^{2,\delta,R}$.

We then claim that

$$
\mathbb{E} \int_{\mathbb{R}^d} \left(\widetilde{G}(x, \widetilde{m}_T^1) - \widetilde{G}(x, \widetilde{m}_T^2) \right) (\widetilde{m}_T^{1,\delta,R}(x) - \widetilde{m}_T^{2,\delta,R}(x)) dx
$$

$$
+ \mathbb{E} \int_0^T \int_{\mathbb{R}^d} \left(\widetilde{F}_t(x, \widetilde{m}_t^1) - \widetilde{F}_t(x, \widetilde{m}_t^2) \right) (\widetilde{m}_t^{1,\delta,R}(x) - \widetilde{m}_t^{2,\delta,R}(x)) dx dt
$$

$$
+ \mathbb{E} \int_0^T \int_{\mathbb{R}^d} \rho_\delta(x) * D(\eta_R(\widetilde{u}_t^1 - \widetilde{u}_t^2))(x) \Big[\widetilde{m}_t^1(x) D_p \widetilde{H}_t(x, D\widetilde{u}_t^1(x))
$$

(4.37)
$$
- \widetilde{m}_t^2(x) D_p \widetilde{H}_t(x, D\widetilde{u}_t^2(x)) \Big] dx dt
$$

$$
- \mathbb{E} \int_0^T \int_{\mathbb{R}^d} \left(\widetilde{H}_t(x, D\widetilde{u}_t^1(x)) - \widetilde{H}_t(x, D\widetilde{u}_t^2(x)) \right) (\widetilde{m}_t^{1,\delta,R}(x) - \widetilde{m}_t^{2,\delta,R}(x)) dx dt
$$

$$
= \mathbb{E} \int_0^T \int_{\mathbb{R}^d} (\widetilde{u}_t^1(x) - \widetilde{u}_t^2(x))(\alpha_t^{\delta,R}(x) + \beta_t^{\delta,R}(x)) dx dt.
$$

We omit the full details of the argument, which is similar to [10, Theorem 3.3]. In short, for a fine partition $\{0 = t_0 < t_1 < t_2 < \cdots < t_N = T\}$ of $[0,T]$, the increments

$$
\int_{\mathbb{R}^d} (\widetilde{u}_{t_{i+1}}^1(x) - \widetilde{u}_{t_{i+1}}^2(x))(\widetilde{m}_{t_{i+1}}^{1,\delta,R}(x) - \widetilde{m}_{t_{i+1}}^{2,\delta,R}(x)) dx
$$

$$
- \int_{\mathbb{R}^d} (\widetilde{u}_{t_i}^1(x) - \widetilde{u}_{t_i}^2(x))(\widetilde{m}_{t_i}^{1,\delta,R}(x) - \widetilde{m}_{t_i}^{2,\delta,R}(x)) dx
$$

are decomposed, using Definition 3.1 and (4.36); note that $\widetilde{m}^{1,\delta,R}, \widetilde{m}^{2,\delta,R}$ are smooth and compactly supported in the state variable. Taking the expectation, exploiting the martingale property of M^1 and M^2, and sending the size of the partition to 0 then yields (4.37).

We have $\widetilde{m}^1, \widetilde{m}^2 \in L^\infty(\Omega, (L^\infty \cap L^1)([0,T] \times \mathbb{R}^d))$ and $D\widetilde{u}^1, D\widetilde{u}^2 \in L^\infty(\Omega \times [0,T] \times \mathbb{R}^d)$, as well as $\|D\eta_R\|_\infty \leq CR^{-1}$ for some universal constant $C > 0$, and so, sending first $\delta \to 0$ and then $R \to \infty$, the left-hand side of (4.37) becomes

$$
\mathbb{E} \int_{\mathbb{R}^d} \left(\widetilde{G}(x, \widetilde{m}_T^1) - \widetilde{G}(x, \widetilde{m}_T^2) \right) (\widetilde{m}_T^1(x) - \widetilde{m}_T^2(x)) dx
$$

$$
+ \mathbb{E} \int_0^T \int_{\mathbb{R}^d} \left(\widetilde{F}_t(x, \widetilde{m}_t^1) - \widetilde{F}_t(x, \widetilde{m}_t^2) \right) (\widetilde{m}_t^1(x) - \widetilde{m}_t^2(x)) dx dt
$$

$$
+ \mathbb{E} \int_0^T \int_{\mathbb{R}^d} (D\widetilde{u}_t^1(x) - D\widetilde{u}_t^2(x)) \Big[\widetilde{m}_t^1(x) D_p \widetilde{H}_t(x, D\widetilde{u}_t^1(x))
$$

$$
- \widetilde{m}_t^2(x) D_p \widetilde{H}_t(x, D\widetilde{u}_t^2(x)) \Big] dx dt
$$

$$
- \mathbb{E} \int_0^T \int_{\mathbb{R}^d} \left(\widetilde{H}_t(x, D\widetilde{u}_t^1(x)) - \widetilde{H}_t(x, D\widetilde{u}_t^2(x)) \right) (\widetilde{m}_t^1(x) - \widetilde{m}_t^2(x)) dx dt
$$

$$
\geq \mathbb{E} \int_{\mathbb{R}^d} \left(\widetilde{G}(x, \widetilde{m}_T^1) - \widetilde{G}(x, \widetilde{m}_T^2) \right) (\widetilde{m}_T^1(x) - \widetilde{m}_T^2(x)) dx
$$

$$
+ \mathbb{E} \int_0^T \int_{\mathbb{R}^d} \left(\widetilde{F}_t(x, \widetilde{m}_t^1) - \widetilde{F}_t(x, \widetilde{m}_t^2) \right) (\widetilde{m}_t^1(x) - \widetilde{m}_t^2(x)) dx dt,
$$

where the inequality follows from the convexity of $\widetilde{H}$.

Define $\mu := D^2(\widetilde{u}_t^1 - \widetilde{u}_t^2)$, which belongs to $L^\infty([0,T] \times \Omega, \mathcal{M}_{\mathrm{loc}}(\mathbb{R}^d) \cap W^{-1,\infty}(\mathbb{R}^d))$ in view of Lemma 2.1(c) and the uniform Lipschitz and semiconcavity constants for $\widetilde{u}^1$ and $\widetilde{u}^2$. Then

$$\int_0^T \int_{\mathbb{R}^d} (\widetilde{u}_t^1(x) - \widetilde{u}_t^2(x)) \alpha_t^{\delta,R}(x) dx dt$$

$$= \int_0^T \left[\mathrm{tr} \left\langle \mu_t, \eta_R \int_{\mathbb{R}^d} (\widetilde{a}_t(y) - \widetilde{a}_t(\cdot))(\widetilde{m}_t^1(y) - \widetilde{m}_t^2(y))\rho_\delta(\cdot - y) dy \right\rangle \right.$$

$$- \int_{\mathbb{R}^d} (D\widetilde{u}_t^1(x) - D\widetilde{u}_t^2(x))$$

$$\cdot \int_{\mathbb{R}^d} (\widetilde{a}_t(y) - \widetilde{a}_t(x))(\widetilde{m}_t^1(y) - \widetilde{m}_t^2(y))\rho_\delta(x-y) dy D\eta_R(x) dx$$

$$+ \int_{\mathbb{R}^d} (\widetilde{u}_t^1(x) - \widetilde{u}_t^2(x))$$

$$\left. \cdot \mathrm{tr} \left[D^2 \eta_R(x) \int_{\mathbb{R}^d} (\widetilde{a}_t(y) - \widetilde{a}_t(x))(\widetilde{m}_t^1(y) - \widetilde{m}_t^2(y))\rho_\delta(x-y) dy \right] dx \right] dt.$$

For fixed $R > 0$, the continuous function

$$x \mapsto \eta_R(x) \int_{\mathbb{R}^d} (\widetilde{a}_t(y) - \widetilde{a}_t(x))(\widetilde{m}_t^1(y) - \widetilde{m}_t^2(y))\rho_\delta(x-y) dy$$

is supported in B_R, independently of δ, and, as $\delta \to 0$, it converges uniformly to 0. It follows that the pairing of μ with this function vanishes as $\delta \to 0$, and, therefore,

$$\lim_{\delta \to 0} \int_0^T \int_{\mathbb{R}^d} (\widetilde{u}_t^1(x) - \widetilde{u}_t^2(x)) \alpha_t^{\delta,R}(x) dx = 0.$$

We also have

$$\int_{\mathbb{R}^d} (\widetilde{u}_t^1(x) - \widetilde{u}_t^2(x)) \beta_t^{\delta,R}(x) dx$$

$$= \int_{\mathbb{R}^d} \left(\widetilde{m}_t^{1,\delta}(x) - \widetilde{m}_t^{2,\delta}(x) \right) \widetilde{a}_t(x) D\eta_R(x) \cdot (D\widetilde{u}_t^1(x) - D\widetilde{u}_t^2(x)) dx,$$

and so, for some constant $C > 0$ independent of δ and R, with probability one,

$$\left| \int_0^T \int_{\mathbb{R}^d} (\widetilde{u}_t^1(x) - \widetilde{u}_t^2(x)) \beta_t^{\delta,R}(x) dx dt \right| \le \frac{C}{R}.$$

We conclude, upon sending first $\delta \to 0$ and then $R \to 0$ on both sides of (4.37), that

$$\mathbb{E} \int_{\mathbb{R}^d} \left(\widetilde{G}(x, \widetilde{m}_T^1) - \widetilde{G}(x, \widetilde{m}_T^2) \right) (\widetilde{m}_T^1(x) - \widetilde{m}_T^2(x)) dx$$

$$+ \mathbb{E} \int_0^T \int_{\mathbb{R}^d} \left(\widetilde{F}_t(x, \widetilde{m}_t^1) - \widetilde{F}_t(x, \widetilde{m}_t^2) \right) (\widetilde{m}_t^1(x) - \widetilde{m}_t^2(x)) dx dt \le 0.$$

The strict monotonicity stated in (4.33) for G and F implies that $\widetilde{G}$ and $\widetilde{F}$ are strictly monotone, and so we conclude that $\widetilde{m}^1 = \widetilde{m}^2$. The proof is finished in view of Theorem 3.2, which implies that $\widetilde{u}^1 = \widetilde{u}^2$ and $M^1 = M^2$. $\qquad\square$

REFERENCES

[1] BAYRAKTAR, E., CECCHIN, A., COHEN, A. and DELARUE, F. (2021). Finite state mean field games with Wright-Fisher common noise. *J. Math. Pures Appl. (9)* **147** 98–162. https://doi.org/10.1016/j.matpur.2021.01.003 MR4213680

[2] BERTUCCI, C. (2021). Monotone solutions for mean field games master equations: finite state space and optimal stopping. *J. Éc. polytech. Math.* **8** 1099–1132. https://doi.org/10.5802/jep.167 MR4275225

[3] BERTUCCI, C. (2023). Monotone solutions for mean field games master equations: continuous state space and common noise. *Comm. Partial Differential Equations* **48** 1245–1285. https://doi.org/10.1080/03605302.2023.2276564 MR4687277

[4] BERTUCCI, C., LASRY, J.-M. and LIONS, P.-L. (2019). Some remarks on mean field games. *Comm. Partial Differential Equations* **44** 205–227. https://doi.org/10.1080/03605302.2018.1542438 MR3941633

[5] BERTUCCI, C., LASRY, J.-M. and LIONS, P.-L. (2021). Master equation for the finite state space planning problem. *Arch. Ration. Mech. Anal.* **242** 327–342. https://doi.org/10.1007/s00205-021-01687-8 MR4302761

[6] CANNARSA, P. and SINESTRARI, C. (2004). *Semiconcave functions, Hamilton-Jacobi equations, and optimal control*. *Progress in Nonlinear Differential Equations and their Applications* **58**. Birkhäuser Boston, Inc., Boston, MA. MR2041617

[7] CARDALIAGUET, P., CIRANT, M. and PORRETTA, A. (2023). Splitting methods and short time existence for the master equations in mean field games. *J. Eur. Math. Soc. (JEMS)* **25** 1823–1918. https://doi.org/10.4171/jems/1227 MR4592861

[8] CARDALIAGUET, P., DELARUE, F., LASRY, J.-M. and LIONS, P.-L. (2019). *The master equation and the convergence problem in mean field games*. *Annals of Mathematics Studies* **201**. Princeton University Press, Princeton, NJ. https://doi.org/10.2307/j.ctvckq7qf MR3967062

[9] CARDALIAGUET, P. and SOUGANIDIS, P. (2022). Monotone solutions of the master equation for mean field games with idiosyncratic noise. *SIAM J. Math. Anal.* **54** 4198–4237. https://doi.org/10.1137/21M1450008 MR4451309

[10] CARDALIAGUET, P. and SOUGANIDIS, P. E. (2022). On first order mean field game systems with a common noise. *Ann. Appl. Probab.* **32** 2289–2326. https://doi.org/10.1214/21-aap1734 MR4430014

[11] CARMONA, R. and DELARUE, F. (2018). *Probabilistic theory of mean field games with applications. II. Probability Theory and Stochastic Modelling* **84**. Springer, Cham Mean field games with common noise and master equations. MR3753660

[12] CARMONA, R., DELARUE, F. and LACKER, D. (2016). Mean field games with common noise. *Ann. Probab.* **44** 3740–3803. https://doi.org/10.1214/15-AOP1060 MR3572323

[13] CATELLIER, R. and GUBINELLI, M. (2016). Averaging along irregular curves and regularisation of ODEs. *Stochastic Process. Appl.* **126** 2323–2366. https://doi.org/10.1016/j.spa.2016.02.002 MR3505229

[14] DAVIE, A. M. (2007). Uniqueness of solutions of stochastic differential equations. *Int. Math. Res. Not. IMRN* **24** Art. ID rnm124, 26. https://doi.org/10.1093/imrn/rnm124 MR2377011

[15] DJETE, M. F. (2023). Large population games with interactions through controls and common noise: convergence results and equivalence between open-loop and closed-loop controls. *ESAIM Control Optim. Calc. Var.* **29** Paper No. 39, 42. https://doi.org/10.1051/cocv/2023005 MR4598796

[16] DOUGLIS, A. (1961). The continuous dependence of generalized solutions of non-linear partial differential equations upon initial data. *Comm. Pure Appl. Math.* **14** 267–284. https://doi.org/10.1002/cpa.3160140307 MR139848

[17] EVANS, L. C. (2010). *Partial differential equations*, second ed. *Graduate Studies in Mathematics* **19**. American Mathematical Society, Providence, RI. https://doi.org/10.1090/gsm/019 MR2597943

[18] EVANS, L. C. and GARIEPY, R. F. (2015). *Measure theory and fine properties of functions*, revised ed. *Textbooks in Mathematics*. CRC Press, Boca Raton, FL. MR3409135

[19] FLANDOLI, F., GUBINELLI, M. and PRIOLA, E. (2010). Well-posedness of the transport equation by stochastic perturbation. *Invent. Math.* **180** 1–53. https://doi.org/10.1007/s00222-009-0224-4 MR2593276

[20] FLEMING, W. H. (1969). The Cauchy problem for a nonlinear first order partial differential equation. *J. Differential Equations* **5** 515–530. https://doi.org/10.1016/0022-0396(69)90091-6 MR235269

[21] GANGBO, W., MÉSZÁROS, A. R., MOU, C. and ZHANG, J. Mean Field Games Master Equations with Non-separable Hamiltonians and Displacement Monotonicity. Preprint, arXiv:2101.12362 [math.AP].

[22] ISHII, H. and LIONS, P. L. (1990). Viscosity solutions of fully nonlinear second-order elliptic partial differential equations. *J. Differential Equations* **83** 26–78. https://doi.org/10.1016/0022-0396(90)90068-Z MR1031377

[23] KRUŽKOV, S. N. (1960). The Cauchy problem in the large for certain non-linear first order differential equations. *Soviet Math. Dokl.* **1** 474–477. MR0121575

[24] LASRY, J.-M. and LIONS, P.-L. (2006). Jeux à champ moyen. I. Le cas stationnaire. *C. R. Math. Acad. Sci. Paris* **343** 619–625. https://doi.org/10.1016/j.crma.2006.09.019 MR2269875

[25] LASRY, J.-M. and LIONS, P.-L. (2006). Jeux à champ moyen. II. Horizon fini et contrôle optimal. *C. R. Math. Acad. Sci. Paris* **343** 679–684. https://doi.org/10.1016/j.crma.2006.09.018 MR2271747

[26] LASRY, J.-M. and LIONS, P.-L. (2007). Mean field games. *Jpn. J. Math.* **2** 229–260. https://doi.org/10.1007/s11537-007-0657-8 MR2295621

[27] LIONS, P. L. Courses at the Collège de France.

[28] LIONS, P. L. (1983). Optimal control of diffusion processes and Hamilton-Jacobi-Bellman equations. III. Regularity of the optimal cost function. In *Nonlinear partial differential equations and their applications. Collège de France seminar, Vol. V (Paris, 1981/1982). Res. Notes in Math.* **93** 95–205. Pitman, Boston, MA. MR725360

[29] LIONS, P. L. (1983). Optimal control of diffusion processes and Hamilton-Jacobi-Bellman equations. II. Viscosity solutions and uniqueness. *Comm. Partial Differential Equations* **8** 1229–1276. https://doi.org/10.1080/03605308308820301 MR709162

[30] MOU, C. and ZHANG, J. (2024). Wellposedness of second order master equations for mean field games with nonsmooth data. *Mem. Amer. Math. Soc.* **302** v+86. https://doi.org/10.1090/memo/1515 MR4813035

[31] PENG, S. G. (1992). Stochastic Hamilton-Jacobi-Bellman equations. *SIAM J. Control Optim.* **30** 284–304. https://doi.org/10.1137/0330018 MR1149069

[32] QIU, J. (2018). Viscosity solutions of stochastic Hamilton-Jacobi-Bellman equations. *SIAM J. Control Optim.* **56** 3708–3730. https://doi.org/10.1137/17M1148232 MR3864678

[33] QIU, J. and WEI, W. (2019). Uniqueness of viscosity solutions of stochastic Hamilton-Jacobi equations. *Acta Math. Sci. Ser. B (Engl. Ed.)* **39** 857–873. https://doi.org/10.1007/s10473-019-0314-3 MR4066509

[34] YAMADA, T. and WATANABE, S. (1971). On the uniqueness of solutions of stochastic differential equations. *J. Math. Kyoto Univ.* **11** 155–167. https://doi.org/10.1215/kjm/1250523691 MR278420