Available online at www.sciencedirect.com
ScienceDirect
J. Differential Equations 269 (2020) 4503–4543
Journal of Differential Equations
www.elsevier.com/locate/jde

Non-coercive first order Mean Field Games

Paola Mannucci [a,*], Claudio Marchi [b], Carlo Mariconda [a], Nicoletta Tchou [c]

[a] Dipartimento di Matematica "Tullio Levi-Civita", Università di Padova, I-35121 Padova, Italy
[b] Dipartimento di Ingegneria dell'Informazione, Università di Padova, I-35131 Padova, Italy
[c] Université de Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France

Received 10 January 2019; revised 25 November 2019; accepted 22 March 2020
Available online 31 March 2020

Abstract

We study first order evolutive Mean Field Games where the Hamiltonian are non-coercive. This situation occurs, for instance, when some directions are "forbidden" to the generic player at some points. We establish the existence of a weak solution of the system via a vanishing viscosity method and, mainly, we prove that the evolution of the population's density is the push-forward of the initial density through the flow characterized almost everywhere by the optimal trajectories of the control problem underlying the Hamilton-Jacobi equation. As preliminary steps, we need to prove that the optimal trajectories for the control problem are unique (at least for a.e. starting points) and that the corresponding unique optimal control has a feedback expression in terms of the intrinsic gradient of the value function.

MSC: 35F50; 35Q91; 49K20; 49L25

Keywords: Mean Field Games; First order Hamilton-Jacobi equations; Continuity equation; Non-coercive Hamiltonian; Degenerate optimal control problem

* Corresponding author.
E-mail addresses: mannucci@math.unipd.it (P. Mannucci), claudio.marchi@unipd.it (C. Marchi), carlo.mariconda@unipd.it (C. Mariconda), nicoletta.tchou@univ-rennes1.fr (N. Tchou).

https://doi.org/10.1016/j.jde.2020.03.035
0022-0396/

1. Introduction

In this paper we study the following Mean Field Game (briefly, MFG)

$$\begin{cases} (i) & -\partial_t u + H(x, Du) = F[m(t)](x) & \text{in } \mathbb{R}^2 \times (0, T) \\ (ii) & \partial_t m - \operatorname{div}(m\, \partial_p H(x, Du)) = 0 & \text{in } \mathbb{R}^2 \times (0, T) \\ (iii) & m(x, 0) = m_0(x), u(x, T) = G[m(T)](x) & \text{on } \mathbb{R}^2, \end{cases} \tag{1.1}$$

where, if $p = (p_1, p_2)$ and $x = (x_1, x_2)$, the functions $H(x, p)$ is

$$H(x, p) = \frac{1}{2}(p_1^2 + h^2(x_1) p_2^2) \tag{1.2}$$

where $h(x_1)$ is a regular bounded function *possibly vanishing* and that F and G are strongly regularizing coupling operators (see assumptions (H1) – (H4) below).

These MFG systems arise when the dynamics of the generic player are deterministic and, when h vanishes, may have a "forbidden" direction; actually, if the evolution of the whole population's distribution m is given, each agent wants to choose the control $\alpha = (\alpha_1, \alpha_2)$ in $L^2([t, T]; \mathbb{R}^2)$ in order to minimize the cost

$$\int_t^T \left[\frac{1}{2}|\alpha(\tau)|^2 + F[m(\tau)](x(\tau))\right] d\tau + G[m(T)](x(T)) \tag{1.3}$$

where, in $[t, T]$, its dynamics $x(\cdot)$ are governed by

$$\begin{cases} x_1'(s) = \alpha_1(s) \\ x_2'(s) = h(x_1(s))\alpha_2(s) \end{cases} \tag{1.4}$$

with $x_1(t) = x_1$ and $x_2(t) = x_2$. We see that the direction along x_2 is forbidden when $h(x_1)$ has zero value. These kinds of problems are called of "Grushin type" (see [28] or Example 1.1 below).

In the present paper we focus our attention to this two dimensional model because it already contains all the main technical issues, however in Section 5 we will consider a generalization to the d-dimensional case where the dynamics are governed by

$$x'(s) = \alpha(s) B^T(x(s)), \tag{1.5}$$

and $B(x)$ is a triangular matrix with a particular structure (see (5.3)). As a matter of fact the structure of the degenerate dynamics will play an essential role in our results because it is sufficient for deriving several properties of optimal trajectories mainly their regularity and a uniform L^∞ estimate of the optimal control laws.

Let us recall that the MFG theory studies Nash equilibria in games with a huge number of ("infinitely many") rational and indistinguishable agents. This theory started with the pioneering papers by Lasry and Lions [24–26] and by Huang, Malhamé and Caines [21]. A detailed description of the achievements obtained in these years goes beyond the scope of this paper; we just refer the reader to the monographs [1,12,7,19,20].

As far as we know, degenerate MFG systems have been poorly investigated up to now. Dragoni and Feleqi [18] studied a second order (stationary) system where the principal part of the operator fulfills the Hörmander condition; moreover, for variational MFG, Cardaliaguet, Graber, Porretta and Tonon [14] tackled degenerate second order systems with coercive (and convex as well) first order operators. Hence, these results cannot be directly applied to the non-coercive problem (1.1).

The purpose of this paper is to extend to our degenerate case two properties of MFG systems with coercive Hamiltonians (see [12]): the first one is to prove the existence of a solution of (1.1), while the second, and main result is the expression of the evolution of the population's density m as the push-forward of the distribution at the initial time through the flow characterized almost everywhere by the optimal trajectories of the control problem underlying the Hamilton-Jacobi equation. Roughly speaking, as in the Lagrangian approach for MFG (see [6,9]), this property means that for a.e. starting positions, the agents follow the optimal trajectories associated to the Hamilton-Jacobi equation.

In order to establish the representation formula for m, we shall follow some ideas of P-L Lions in the lectures at Collège de France (see [12]) but, since H is non-coercive, we have to apply also some techniques of [15,13] and the superposition principle [3]. To this end we have to study carefully the behaviour of the optimal trajectories of the control problem associated to the Hamilton-Jacobi equation (1.1)-(i) especially their uniqueness and their regularity. Crucial points will be the application of the Pontryagin maximum principle and the statement of Theorem 2.1 on the uniqueness of the optimal trajectory after a rest time. As far as we know this uniqueness property has never been tackled before for this kind of degenerate dynamics and, in our opinion, it may have interest in itself.

We point out that our approach could be applied to other first order "degenerate" MFG systems but it is essential to prove some uniqueness properties of optimal trajectories in a set of starting points of full measure. In general this set depends on the semiconcavity properties of u (see Definition 2.3), as in the classical setting, and on the degeneracy of the dynamics.

We now list our notations and the assumptions, we give the definition of (weak) solution to system (1.1) and we state the existence result for system (1.1).

Notations and Assumptions. For $x = (x_1, x_2) \in \mathbb{R}^2$, $\phi : \mathbb{R}^2 \to \mathbb{R}$ and $\Phi : \mathbb{R}^2 \to \mathbb{R}^2$ differentiable, we set: $D_{\mathcal{G}}\phi(x) := (\partial_{x_1}\phi(x), h(x_1)\partial_{x_2}\phi(x))$ and $\mathrm{div}_{\mathcal{G}}\,\Phi(x) := \partial_{x_1}\Phi_1(x) + h(x_1)\partial_{x_2}\Phi_2(x)$. We denote by $\mathcal{P}_1$ the space of Borel probability measures on $\mathbb{R}^2$ with finite first order moment, endowed with the Kantorovich-Rubinstein distance $\mathbf{d}_1$. We denote $C^2(\mathbb{R}^2)$ the space of functions with continuous second order derivatives endowed with the norm

$$\|f\|_{C^2} := \sup_{x\in\mathbb{R}^2} [|f(x)| + |Df(x)| + |D^2 f(x)|].$$

Throughout this paper, unless otherwise explicitly stated, we shall require the following hypotheses:

(H1) the functions F and G are real-valued function, continuous on $\mathcal{P}_1 \times \mathbb{R}^2$;

(H2) the map $m \to F[m](\cdot)$ is Lipschitz continuous from $\mathcal{P}_1$ to $C^2(\mathbb{R}^2)$; moreover, there exists $C \in \mathbb{R}$ such that

$$\|F[m](\cdot)\|_{C^2}, \|G[m](\cdot)\|_{C^2} \le C, \qquad \forall m \in \mathcal{P}_1;$$

(H3) the function $h : \mathbb{R} \to \mathbb{R}$ is $C^2(\mathbb{R})$ with $\|h\|_{C^2} \le C$ and $\mathcal{Z} := \{z \in \mathbb{R} : h(z) = 0\}$ has null measure;

(H4) the initial distribution m_0 has a compactly supported density (that we still denote by m_0, with a slight abuse of notation), $m_0 \in C^{2,\delta}(\mathbb{R}^2)$, for a $\delta \in (0,1)$.

Example 1.1. Easy examples of h are $h(x_1) = \sin(x_1)$ or $h(x_1) = \frac{x_1}{\sqrt{1+x_1^2}}$, (see [28] where the term $h(x_1) = \frac{x_1}{\sqrt{1+x_1^2}}$ is introduced as a degenerate diffusion term).

We now introduce our definition of solution of the MFG system (1.1) and state the main result concerning its existence.

Definition 1.1. The pair (u,m) is a solution of system (1.1) if:

1) $(u,m) \in W^{1,\infty}(\mathbb{R}^2 \times [0,T]) \times C^0([0,T]; \mathcal{P}_1(\mathbb{R}^2))$ and for all $t \in [0,T]$, $m(t)$ is absolutely continuous w.r.t. the Lebesgue measure on $\mathbb{R}^2$. Let $m(\cdot,t)$ denote the density of $m(t)$. The function $(x,t) \mapsto m(x,t)$ is bounded;
2) Equation (1.1)-(i) is satisfied by u in the viscosity sense;
3) Equation (1.1)-(ii) is satisfied by m in the sense of distributions.

Here below we state the main result of this paper.

Theorem 1.1. *Under the above assumptions:*

1. *System* (1.1) *has a solution* (u,m) *in the sense of Definition 1.1,*
2. *m is the push-forward of m_0 through the characteristic flow*

$$\begin{cases} x_1'(s) = -u_{x_1}(x(s),s), & x_1(0) = x_1, \\ x_2'(s) = -h^2(x_1(s))u_{x_2}(x(s),s), & x_2(0) = x_2. \end{cases} \tag{1.6}$$

Remark 1.1. Uniqueness holds under classical hypothesis on the monotonicity of F and G as in [12].

This paper is organized as follows. Section 2 is devoted to the study of optimal trajectories in this degenerate case. We will establish a crucial point of the paper: a uniqueness property of the optimal trajectory of the associated control problem. Moreover still in this section we will find some properties of the solution u of the Hamilton-Jacobi equation (1.1)-(i) with fixed m: we will prove that u is Lipschitz continuous in (x,t) and semiconcave in x. In Section 3 we study the continuity equation (1.1)-(ii) where u is the solution of the Hamilton-Jacobi equation found in the previous section. Section 4 is devoted to the proof of the Theorem 1.1. Section 5 is devoted to illustrate the d-dimensional case; since the techniques are very similar to those used in the previous sections we will just show the new issues. Finally, the Appendix splits into three parts: in the first one, we give some results on the concatenation of optimal trajectories and the Dynamic Programming Principle while in the last two parts we introduce the notion of the $\mathcal{G}$-differentiability and we prove the main properties on the $\mathcal{G}$-differentials which will be used along the paper. Indeed in this degenerate case the results on differentials and semiconcavity established in [11] cannot be directly applied.

2. Formulation of the optimal control problem

For every $0 \leq t \leq T$ and $x := (x_1, x_2) \in \mathbb{R}^2$ we consider the following optimal control problem, where the functions f, g, h satisfy the Hypothesis 2.1 here below.

Definition 2.1 *(Optimal Control Problem (OC)).*

$$\text{Minimize } J_t(x(\cdot), \alpha) := \int_t^T \frac{1}{2}|\alpha(s)|^2 + f(x(s), s)\, ds + g(x(T)) \tag{2.1}$$

subject to $(x(\cdot), \alpha) \in \mathcal{A}(x, t)$, where

$$\mathcal{A}(x,t) := \left\{ (x(\cdot), \alpha(\cdot)) \in AC([t,T];\mathbb{R}^2) \times L^2([t,T];\mathbb{R}^2) : (1.4) \text{ holds a.e. with } x(t) = x \right\}. \tag{2.2}$$

A pair $(x(\cdot), \alpha)$ in $\mathcal{A}(x, t)$ is said to be admissible. We say that x^* is an optimal trajectory if there is a control α^* such that (x^*, α^*) is optimal for (OC). Also, we shall refer to the system (1.4) as to the dynamics of the optimal control problem (OC).

In what follows, the functions f, g and h satisfy the following conditions.

Hypothesis 2.1.

(i) $f \in C^0([0, T], C^2(\mathbb{R}^2))$ and there exists a constant C such that

$$\|f(\cdot, t)\|_{C^2(\mathbb{R}^2)} + \|g\|_{C^2(\mathbb{R}^2)} + \|h\|_{C^2(\mathbb{R})} \leq C, \qquad \forall t \in [0, T].$$

(ii) The set $\mathcal{Z} = \{z \in \mathbb{R} \,:\, h(z) = 0\}$ has null measure.

Condition (ii) will play a crucial role to prove some stationary condition (see Lemma 2.1) and the uniqueness of the optimal trajectory after a rest time (see Theorem 2.1).

Remark 2.1. Notice that, given a control law $\alpha \in L^2([t, T]; \mathbb{R}^2)$, the Hypothesis 2.1 on h implies that, given the initial point x, there is a *unique* trajectory $x(\cdot)$ such that $(x(\cdot), \alpha) \in \mathcal{A}(x, t)$.

Remark 2.2 *(Existence of optimal solutions).* Hypothesis 2.1-(i) ensures that the optimal control problem (OC) admits a solution (x^*, α^*).

Definition 2.2. The value function for the cost J_t defined in (2.1) is

$$u(x, t) := \inf\{J_t(x(\cdot), \alpha) : (x(\cdot), \alpha) \in \mathcal{A}(x, t)\}. \tag{2.3}$$

An optimal pair $(x^*(\cdot), \alpha^*)$ for the control problem (OC) in Definition 2.1 is also said to be optimal for $u(x, t)$.

2.1. Necessary conditions and regularity for the optimal trajectories

The application of the Maximum Principle (see [16, Theorem 22.17]) yields the following necessary conditions.

Proposition 2.1 *(Necessary conditions for optimality). Let* (x^*, α^*) *be optimal for* (OC). *There exists an arc* $p \in AC([t, T]; \mathbb{R}^2)$*, hereafter called the costate, such that*

1. *The pair* (α^*, p) *satisfies the adjoint equations: for a.e.* $s \in [t, T]$

$$p_1' = -p_2 h'(x_1^*)\alpha_2^* + f_{x_1}(x^*, s) \tag{2.4}$$

$$p_2' = f_{x_2}(x^*, s), \tag{2.5}$$

the transversality condition

$$-p(T) = Dg(x^*(T)) \tag{2.6}$$

together with the maximum condition

$$\max_{\alpha=(\alpha_1,\alpha_2)\in\mathbb{R}^2} p_1(s)\alpha_1 + p_2(s)h(x_1^*(s))\alpha_2 - \frac{|\alpha|^2}{2} =$$

$$= p_1(s)\alpha_1^*(s) + p_2(s)h(x_1^*(s))\alpha_2^*(s) - \frac{|\alpha^*(s)|^2}{2} \text{ a.e. } s \in [t, T]. \tag{2.7}$$

2. *The optimal control* α^* *is given by*

$$\begin{cases} \alpha_1^* = p_1 \\ \alpha_2^* = p_2 h(x_1^*) \end{cases} \quad \text{a.e on } [t, T]. \tag{2.8}$$

3. *The pair* (x^*, p) *satisfies the system of differential equations: for a.e.* $s \in [t, T]$

$$x_1' = p_1 \tag{2.9}$$

$$x_2' = h^2(x_1)p_2 \tag{2.10}$$

$$p_1' = -p_2^2 h'(x_1)h(x_1) + f_{x_1}(x, s) \tag{2.11}$$

$$p_2' = f_{x_2}(x, s) \tag{2.12}$$

with the mixed boundary conditions $x^*(t) = x,\ p(T) = -Dg(x^*(T))$.

Proof. 1. Hypothesis 2.1 -(i) ensures the validity of the assumptions of the Maximum Principle [16, Theorem 22.17] with the Hamiltonian

$$H^\eta(s, x_1, x_2, p_1, p_2, \alpha_1, \alpha_2) = p_1\alpha_1 + p_2 h(x_1)\alpha_2 - \eta\left(\frac{1}{2}|\alpha|^2 + f(x, s)\right).$$

Since the endpoint is free, [16, Corollary 22.3] implies that the deduced necessary conditions hold in normal form (i.e., with $\eta = 1$): the claim follows directly.

2. The maximum condition (2.7) implies that

$$D_\alpha \left(p_1(s)\alpha_1 + p_2(s)h(x_1^*(s))\alpha_2 - \frac{|\alpha|^2}{2} \right)_{\alpha=\alpha^*} = 0 \quad \text{a.e. } s \in [t,T]$$

from which we get (2.8).

3. Conditions (2.9) – (2.10) follow directly from the dynamics (1.4) replacing α_1^*, α_2^* by means of (2.8). Condition (2.11) follows similarly from (2.4), whereas (2.12) coincides with (2.5). □

Let us emphasize that the next corollary establishes the regularity of optimal trajectories and an L^∞ estimate of optimal control laws independent of the starting point (x,t). This will be an essential tool to prove the Lipschitz property in Lemma 2.3-(2).

Corollary 2.1 *(Feedback control and regularity). Let (x^*, α^*) be optimal for (OC) starting from (x,t) and p be the related costate as in Proposition 2.1. Then:*

1. *The costate $p = (p_1, p_2)$ is uniquely expressed in terms of x^* for every $s \in [t,T]$ by*

$$\begin{cases} p_1(s) = -g_{x_1}(x^*(T)) - \displaystyle\int_s^T f_{x_1}(x^*,\tau) - p_2^2 h'(x_1^*)h(x_1^*)\,d\tau, \\ p_2(s) = -g_{x_2}(x^*(T)) - \displaystyle\int_s^T f_{x_2}(x^*,\tau)\,d\tau. \end{cases} \tag{2.13}$$

2. *The optimal control $\alpha^* = (\alpha_1^*, \alpha_2^*)$ is a feedback control (i.e., a function of x^*), uniquely expressed for a.e. $s \in [t,T]$ by*

$$\begin{cases} \alpha_1^*(s) = -g_{x_1}(x^*(T)) + \displaystyle\int_T^s f_{x_1}(x^*,\tau) - p_2^2 h'(x_1^*)h(x_1^*)\,d\tau, \\ \alpha_2^*(s) = p_2(s)h(x_1^*(s)). \end{cases} \tag{2.14}$$

3. *The optimal control α^* and the costate p are of class C^1, the optimal trajectory x^* is of class C^2. In particular the equalities (2.8) – (2.14) do hold for every $s \in [t,T]$.*
4. *There is a constant C independent of (x,t) such that $\|p\|_\infty \le C$ and $\|\alpha^*\|_\infty \le C$.*
5. *Assume that, for some $k \in \mathbb{N}$, $h \in C^{k+1}$ and $Df(x,s)$ is of class C^k. Then α^*, p are of class C^{k+1} and x^* is of class C^{k+2}.*

Proof. Point 1 is an immediate consequence of (2.11) – (2.12) together with the endpoint condition $p(T) = -Dg(x^*(T))$. Point 2 follows then directly from (2.8).

3. Since x^* is continuous, the continuity of α^* follows from (2.14). The dynamics (1.4) then yield that $x^* \in C^1$. Again (1.4) gives $x^* \in C^2$. Relations (2.13) and (2.14) imply, respectively, that p and α^* are of class C^1.

4. By Hypothesis 2.1-(i), the system (2.13) allows to easily obtain the boundedness of p_i, $i = 1, 2$ uniformly on (x, t). By (2.14) we get the statement.

5. The relations (2.13) and the C^1-regularity of x^* and p imply that, actually, $p \in C^2$. Therefore, (2.14) gives the C^2-regularity of α^* and, finally, the dynamics (1.4) yield the C^3-regularity of x^*. Further regularity of x^*, α^* and p follows by a standard bootstrap inductive argument. □

2.2. Uniqueness of the trajectories after the initial time

Next Theorem 2.1 implies that the optimal trajectories for $u(x, t)$ do not bifurcate at any time $r > t$ whenever $h(x_1) \neq 0$ (see Corollary 2.2), otherwise they may rest at x in an interval from the initial time t but they do not bifurcate as soon as they leave x.

Theorem 2.1 *(Uniqueness of the optimal trajectory after the rest time). Under Hypothesis 2.1, let x^* be an optimal trajectory for $u(x, t)$.*

1. *Assume that $h(x_1^*(\tau)) \neq 0$ for some $t < \tau < T$. For every $\tau \leq r < T$ there are no other optimal trajectories for $u(x^*(r), r)$ other than x^*, restricted to $[r, T]$.*
2. *Assume that $h(x_1) = 0$. Let t_{x^*} be the rest time for x^* defined by*

$$t_{x^*} := \sup\{r \in [t, T] : x^* \equiv x \text{ on } [t, r]\}.$$

For every $r > t_{x^*}$ there are no optimal trajectories for $u(x^*(r), r)$, other than x^* restricted to $[r, T]$.

The next Lemma 2.1 relates the initial constancy of a trajectory to a stationary condition and is a key argument of the proof of Point 2 of Theorem 2.1.

Lemma 2.1 *(A stationary condition). Assume that $h(x_1) = 0$. Let $x^* = (x_1^*, x_2^*)$ be a trajectory starting from x at time t, and $r \in [t, T]$. Then*

$$x^* \equiv x \text{ on } [t, r] \Leftrightarrow h(x_1^*) \equiv 0 \text{ on } [t, r]. \tag{2.15}$$

Proof. If $h(x_1^*) = 0$ on $[t, r]$ then x_1^* belongs to set of the zeros of h, which has null measure by Hypothesis 2.1. It follows that $x_1^* \equiv x_1$ on $[t, r]$. Moreover, the dynamics (1.4) imply that $x_2^* \equiv x_2$ on $[t, r]$, so that $x^* \equiv x$ on $[t, r]$. The opposite implication is trivial, since $h(x_1) = 0$. □

Remark 2.3. Let us point out that, assuming $h(x_1) = 0$, the stationary trajectories $x^* = x$ on $[t, r]$ are the only singular trajectories in $[t, r]$ of the optimal control problem (OC) (for the classical definition of singular trajectories see [8, Section 2, Def.19] or [10, Def. 2.3]).

Proof of Theorem 2.1. 1. Let $r \in [\tau, T[$ and y^* be optimal for $u(x^*(r), r)$. Point 1 of Proposition 6.1 in the Appendix ensures that the concatenation z^* of x^* with y^* at r is optimal for $u(x, t)$. Let $p := (p_1, p_2), q := (q_1, q_2)$ be the costates associated to $x^* := (x_1^*, x_2^*)$ and, respectively, to $z^* := (z_1^*, z_2^*)$. Both (x^*, p) and (z^*, q) satisfy (2.9) – (2.12) on $[t, T]$. Now, Corollary 2.1 shows that x^* and z^* are of class C^1. Since $x^* = z^*$ on $[t, \tau]$, the fact that $\tau > t$, together with (2.9), imply

$$p_1(\tau) = (x_1^*)'(\tau) = \lim_{s\to\tau^-} (x_1^*)'(s) = \lim_{s\to\tau^-} (z_1^*)'(s) = (z_1^*)'(\tau) = q_1(\tau),$$

whereas (2.10), and the fact that $h(x_1^*(\tau)) \neq 0$ analogously yield

$$p_2(\tau) = \frac{(x_2^*)'(\tau)}{h^2(x_1^*(\tau))} = \frac{(z_2^*)'(\tau)}{h^2(z_1^*(\tau))} = q_2(\tau).$$

Therefore, both (x^*, p) and (z^*, q) are absolutely continuous solutions to the same Cauchy problem on $[t, T]$, with initial data at τ, for the first order differential system (2.9)-(2.12). The regularity assumptions on f, h and Cauchy-Lipschitz Theorem guarantee the uniqueness of the solution. Thus $x^* = z^*$ on $[\tau, T]$, from which we obtain the desired equality $x^* = y^*$ on $[r, T]$.

2. We assume that $t_{x^*} < T$, otherwise the claim is trivial. We deduce from Lemma 2.1 that there is $\tau \in [t_{x^*}, r]$ satisfying $h(x_1^*(\tau)) \neq 0$. Point 1 of Theorem 2.1 yields the conclusion. □

Corollary 2.2. *Let x^* be an optimal trajectory for $u(x, t)$. If $h(x_1) \neq 0$, for every $t < r < T$ there are no other optimal trajectories starting from $x^*(r)$ at time r, other than x^*, restricted to $[r, T]$.*

2.3. The Hamilton-Jacobi equation and the value function of the optimal control problem

The aim of this section is to study the Hamilton-Jacobi equation (1.1)-(i) with m fixed, namely

$$\begin{cases} -\partial_t u + \frac{1}{2}|D_{\mathcal{G}} u|^2 = f(x,t) & \text{in } \mathbb{R}^2 \times (0,T), \\ u(x,T) = g(x) & \text{on } \mathbb{R}^2 \end{cases} \tag{2.16}$$

where $D_{\mathcal{G}} u(x) := (\partial_{x_1} u(x), h(x_1)\partial_{x_2} u(x))$. Under Hypothesis 2.1, we shall prove several regularity properties of the solution (especially Lipschitz continuity and semiconcavity). As a first step, in the next lemma we show that the solution u of (2.16) can be represented as the value function of the control problem (OC) defined in (2.3).

Lemma 2.2. *Under Hypothesis 2.1, the value function u, defined in (2.3), is the unique bounded uniformly continuous viscosity solution to problem (2.16).*

Proof. The Dynamic Programming Principle (stated in Proposition 6.1 in Appendix below) yields that the value function is a solution to problem (2.16). Applying classical results on uniqueness (see, for example, [11, eq. (7.40) and Thm. 7.4.14]), we obtain the statement. Moreover, taking as admissible control the law $\alpha = 0$, from the representation formula (2.3), using the boundedness of f and g, we have $|u(x,t)| \leq C_T$. □

In the following lemma we prove the Lipschitz continuity in both variables x and t of the value function. We shall need this property for instance to prove the a.e. differentiability of $u(x(\cdot), \cdot)$.

Lemma 2.3 *(Lipschitz continuity). Under Hypothesis 2.1, there hold:*

1. *$u(x, t)$ is Lipschitz continuous with respect to the spatial variable x,*
2. *$u(x, t)$ is Lipschitz continuous with respect to the time variable t.*

Proof. In this proof, C_T will denote a constant which may change from line to line but it always depends only on the constants in the assumptions (especially the Lipschitz constants of f and g) and on T.

1. Let t be fixed. We follow the proof of [12, Lemma 4.7]. From Remark 2.2 we know that there exists $\alpha(\cdot)$ optimal control for $u(x,t)$ and $x(\cdot)$ optimal trajectory i.e.:

$$u(x_1,x_2,t)=\int_t^T \frac{1}{2}|\alpha(s)|^2+f(x(s),s)\,ds+g(x(T)). \tag{2.17}$$

From the boundedness of u (established in Lemma 2.2) and our assumptions, there exists a constant C_T such that $\|\alpha\|_{L^2(t,T)}\le C_T$.

We consider the path $x^*(s)$ starting from $y=(y_1,y_2)$, with control α. Hence

$$\begin{aligned}
x_1^*(s)&=y_1+\int_t^s \alpha_1(\tau)\,d\tau=y_1-x_1+x_1(s)\\
x_2^*(s)&=y_2+\int_t^s h(y_1-x_1+x_1(\tau))\alpha_2(\tau)\,d\tau\\
&=y_2-x_2+x_2(s)+\int_t^s h(y_1-x_1+x_1(\tau))\alpha_2(\tau)-h(x_1(\tau))\alpha_2(\tau)\,d\tau.
\end{aligned}$$

Using the Lipschitz continuity of f and h and the boundedness of h we get

$$\begin{aligned}
&f(x^*(s),s)\\
&\le f(x_1(s),x_2(s),s)+L|y_1-x_1|+\\
&\qquad +L\left|y_2-x_2+\int_t^s h(y_1-x_1+x_1(\tau))\alpha_2(\tau)-h(x_1(\tau))\alpha_2(\tau)\,d\tau\right|\\
&\le f(x_1(s),x_2(s),s)+L|y_1-x_1|+L|y_2-x_2|+L'|y_1-x_1|\int_t^s|\alpha_2(\tau)|\,d\tau\\
&\le f(x(s),s)+L|y_1-x_1|+L|y_2-x_2|+L'|y_1-x_1|T^{\frac12}\left(\int_t^s(\alpha_2(s))^2\,ds\right)^{\frac12}.
\end{aligned}$$

By the same calculations for g and substituting equality (2.17) in

$$u(y_1,y_2,t)\le\int_t^T \frac{1}{2}|\alpha(s)|^2+f(x^*(s),s)\,ds+g(x^*(T)),$$

we get

$$u(y_1, y_2, t) \le u(x_1, x_2, t) + C_T(|y_2 - x_2| + |y_1 - x_1|).$$

Reversing the role of x and y we get the result.

2. Thanks to the boundedness of h and the bound of α uniform on (x, t) proved in Corollary 2.1-4, we can follow the same arguments as those in the proof of [12, Lemma 4.7], hence obtaining

$$|x(s) - x| \le C(s - t)\|\alpha\|_\infty \le C_T(s - t). \quad \square$$

In the following lemma we establish the semiconcavity of $u(x, t)$ w.r.t. x; we recall here below the definition of semiconcavity with linear modulus and we refer the reader to the monograph [9] for further properties.

Definition 2.3. Let $u : \mathbb{R}^d \to \mathbb{R}$. We say that u is *semiconcave* (with linear modulus) if there exists a constant $C \ge 0$ such that for all $\lambda \in [0, 1]$,

$$\lambda u(y) + (1 - \lambda)u(x) - 2u(\lambda y + (1 - \lambda)x) \le C\lambda(1 - \lambda)|y - x|^2$$

for all $x, y \in \mathbb{R}^d$.

The semiconcavity of u will be used in the study of the relationship between the regularity of the value function and the uniqueness of the optimal trajectories. It is worth to remark that it is possible to prove that $u(x, t)$ is also semiconcave with respect to the χ-lines associated to the Grushin dynamics, as introduced in [5, Example 2.4], but this does not seem to be useful to our results.

Lemma 2.4 *(Semiconcavity). Under Hypothesis 2.1, the value function u, defined in* (2.3)*, is semiconcave with respect to the variable x.*

Proof. For any $x, y \in \mathbb{R}^2$ and $\lambda \in [0, 1]$, consider $x_\lambda := \lambda x + (1 - \lambda)y$. Let $\alpha(s)$ and $x_\lambda(s)$ be an optimal control and optimal trajectory for $u(x_\lambda, t)$:

$$x_\lambda(s) = (x_{\lambda,1}(s), x_{\lambda,2}(s)) = \left(x_{\lambda,1} + \int_t^s \alpha_1(\tau)\, d\tau,\ x_{\lambda,2} + \int_t^s h(x_{\lambda,1}(\tau))\alpha_2(\tau)\, d\tau\right).$$

Let $x(s)$ and $y(s)$ satisfy (1.4) with initial condition respectively x and y still with the same control α, optimal for $u(x_\lambda, t)$. We have to estimate $\lambda u(x, t) + (1 - \lambda)u(y, t)$ in terms of $u(x_\lambda, t)$. To this end, arguing as in the proof of [12, Lemma 4.7], we have to estimate the terms $\lambda f(x(s), s) + (1 - \lambda)f(y(s), s)$ and $\lambda g(x(T)) + (1 - \lambda)g(y(T))$.

We explicitly provide the calculations for the second component $x_2(s)$ since the calculations for $x_1(s)$ are the same as in [12]. We have

$$
\begin{aligned}
x_2(s) &= x_2 + \int_t^s h(x_1(\tau))\alpha_2(\tau)\,d\tau \\
&= x_2 - x_{\lambda,2} + x_{\lambda,2}(s) + \int_t^s (h(x_1(\tau)) - h(x_{\lambda,1}(\tau)))\alpha_2(\tau)\,d\tau,
\end{aligned}
$$

and analogously for $y_2(s)$. For the sake of brevity we provide the explicit calculations only for f and we omit the analogous ones for g; we write $f(x_1, x_2) := f(x_1, x_2, s)$. We have

$$
\begin{aligned}
&\lambda f(x(s)) + (1-\lambda) f(y(s)) \\
&\quad = \lambda f\left(x_1(s), x_{\lambda,2}(s) + x_2 - x_{\lambda,2} + \int_t^s (h(x_1(\tau)) - h(x_{\lambda,1}(\tau)))\alpha_2(\tau)\,d\tau \right) \\
&\quad + (1-\lambda) f\left(y_1(s), x_{\lambda,2}(s) + y_2 - x_{\lambda,2} + \int_t^s (h(y_1(\tau)) - h(x_{\lambda,1}(\tau)))\alpha_2(\tau)\,d\tau \right).
\end{aligned}
$$

In the Taylor expansion of f centered in $x_\lambda(s)$ the contribution of the first variable can be dealt with as in [12]. Assuming without any loss of generality $x_1 = y_1$, the contribution of the second variable gives

$$
\begin{aligned}
\lambda f(x(s)) + (1-\lambda) f(y(s)) = f(x_\lambda(s)) + \partial_{x_2} f(x_\lambda(s)) \Big(&\lambda(x_2 - x_{\lambda,2}) + (1-\lambda)(y_2 - x_{\lambda,2}) \\
&+ \lambda \int_t^s (h(x_1(\tau)) - h(x_{\lambda,1}(\tau)))\alpha_2(\tau)\,d\tau \\
&+ (1-\lambda) \int_t^s (h(y_1(\tau)) - h(x_{\lambda,1}(\tau)))\alpha_2(\tau)\,d\tau \Big) + R,
\end{aligned}
$$

where R is the error term of the expansion, namely

$$
\begin{aligned}
R = \lambda \frac{\partial^2_{x_2,x_2} f(\xi_1)}{2} &\left(x_2 - x_{\lambda,2} + \int_t^s (h(x_1(\tau)) - h(x_{\lambda,1}(\tau)))\alpha_2(\tau)\,d\tau \right)^2 \\
&+ (1-\lambda) \frac{\partial^2_{x_2,x_2} f(\xi_2)}{2} \left(y_2 - x_{\lambda,2} + \int_t^s (h(y_1(\tau)) - h(x_{\lambda,1}(\tau)))\alpha_2(\tau)\,d\tau \right)^2, \quad (2.18)
\end{aligned}
$$

for suitable $\xi_1, \xi_2 \in \mathbb{R}^2$.

Since $\lambda(x_2 - x_{\lambda,2}) + (1-\lambda)(y_2 - x_{\lambda,2}) = 0$, we get

$$\lambda f(x(s)) + (1-\lambda) f(y(s)) = f(x_\lambda(s)) + \partial_{x_2} f(x_\lambda(s)) \int_t^s I(\tau)\alpha_2(\tau)\, d\tau + R, \tag{2.19}$$

with $I(\tau) := -h(x_{\lambda,1}(\tau)) + \lambda h(x_1(\tau)) + (1-\lambda)h(y_1(\tau))$. Now, our aim is to estimate $I(\tau)$. Since $x_{\lambda,1}(\tau) = \lambda x_1(\tau) + (1-\lambda)y_1(\tau)$, $x_1(\tau) - x_{\lambda,1}(\tau) = (1-\lambda)(x_1 - y_1)$ and $y_1(\tau) - x_{\lambda,1}(\tau) = \lambda(y_1 - x_1)$, the Taylor expansion for h centered in $x_{\lambda,1}(\tau)$ yields

$$I(\tau) = \frac{1}{2}(1-\lambda)\lambda(y_1 - x_1)^2[(1-\lambda)h''(\overline{\xi}) + \lambda h''(\widetilde{\xi})],$$

for suitable $\overline{\xi}, \widetilde{\xi} \in \mathbb{R}$. Our Hypothesis 2.1 entails

$$|I(\tau)| \le (1-\lambda)\lambda C(y_1 - x_1)^2.$$

Replacing the inequality above in (2.19), we obtain

$$\lambda f(x_2(s)) + (1-\lambda) f(y_2(s)) \le f(x_{\lambda,2}(s)) + C^2 T(1-\lambda)\lambda(y_1 - x_1)^2 + R. \tag{2.20}$$

Let us now estimate the error term R in (2.18). We have

$$\begin{aligned}
&\left(x_2 - x_{\lambda,2} + \int_t^s (h(x_1(\tau)) - h(x_{\lambda,1}(\tau)))\alpha_2(\tau)\, d\tau\right)^2 \\
&\quad \le 2(x_2 - x_{\lambda,2})^2 + 2\left(\int_t^s (h(x_1(\tau)) - h(x_{\lambda,1}(\tau)))\alpha_2(\tau)\, d\tau\right)^2 \\
&\quad \le 2(1-\lambda)^2(x_2 - y_2)^2 + 2C(1-\lambda)^2(x_1 - y_1)^2 \le C(1-\lambda)^2|x-y|^2
\end{aligned}$$

and, analogously

$$\left(y_2 - x_{\lambda,2} + \int_t^s (h(y_1(\tau)) - h(x_{\lambda,1}(\tau)))\alpha_2(\tau)\, d\tau\right)^2 \le C\lambda^2|x-y|^2.$$

Then, replacing these two inequalities in (2.18), we infer

$$R \le C(1-\lambda)\lambda|x-y|^2. \tag{2.21}$$

Taking into account (2.21) and (2.20), we get the semiconcavity of u. □

3. The continuity equation

In this section we want to study equation (1.1)-(ii). Since h is independent of x_2, taking account of (1.2), this partial differential equation can be rewritten as

$$\partial_t m - \partial_{x_1}(m\partial_{x_1}u) - h^2(x_1)\partial_{x_2}(m\partial_{x_2}u) = \partial_t m - \mathrm{div}_{\mathcal{G}}(mD_{\mathcal{G}}u) = 0. \tag{3.1}$$

Hence our aim is to study the well posedness of the problem

$$\begin{cases} \partial_t m - \mathrm{div}_{\mathcal{G}}(m\, D_{\mathcal{G}}u) = 0, & \text{in } \mathbb{R}^2 \times (0,T), \\ m(x,0) = m_0(x), & \text{on } \mathbb{R}^2, \end{cases} \tag{3.2}$$

where u is a solution to problem

$$\begin{cases} -\partial_t u + \frac{1}{2}|D_{\mathcal{G}}u|^2 = F[\overline{m}(t)](x) & \text{in } \mathbb{R}^2 \times (0,T), \\ u(x,T) = G[\overline{m}(T)](x), & \text{on } \mathbb{R}^2, \end{cases} \tag{3.3}$$

where the function $\overline{m}$ is fixed in $C^0([0,T],\mathcal{P}_1)$. Note that this problem is equivalent to (2.16) with a fixed $\overline{m}$.

Observe that, by Lemma 2.3-(1), in (3.2) the drift $v = (\partial_{x_1}u, h^2(x_1)\partial_{x_2}u)$ is only bounded; this lack of regularity prevents to apply the standard results (uniqueness, existence and representation formula of m as the push-forward of m_0 through the characteristic flow; e.g., see [3, Proposition 8.1.8]) for drifts which are Lipschitz continuous in x. We shall overcome this difficulty applying the superposition principle [3, Theorem 8.2.1] and proving several results on the optimal trajectories for the control problem stated in Section 2. The superposition principle yields a representation formula of m as the push-forward of some measure on $C^0([0,T],\mathbb{R}^2)$ through the evaluation map e_t, see the proof of Proposition 3.2. In the following theorem, we shall obtain uniqueness, existence and some regularity result for the solution to (3.2).

Theorem 3.1. *Under assumptions* (H1) – (H4)*, for any $\overline{m} \in C^0([0,T],\mathcal{P}_1)$ problem* (3.2) *has a unique bounded solution m in the sense of Definition 1.1. Moreover $m \in L^\infty([0,T],\mathcal{P}_2)$ and it is a Lipschitz continuous map from $[0,T]$ to $\mathcal{P}_1$ with a Lipschitz constant bounded by $\|Du\|_\infty\|h^2\|_\infty$. Moreover, the function m satisfies:*

$$\int_{\mathbb{R}^2} \phi\, dm(t) = \int_{\mathbb{R}^2} \phi(\overline{\gamma}_x(t))\, m_0(x)\, dx \qquad \forall \phi \in C^0_0(\mathbb{R}^2),\ \forall t \in [0,T] \tag{3.4}$$

where, for a.e. $x \in \mathbb{R}^2$, $\overline{\gamma}_x$ is the solution to (1.6)*.*

The proof of Theorem 3.1 is given in the next two subsections which are devoted to the existence result (see Proposition 3.1), and respectively to the uniqueness result by the representation formula (see Proposition 3.2) and to the Lipschitz regularity (see Corollary 3.1).

3.1. Existence of the solution

As in [13, Appendix] (see also [12, Section 4.4]), we now want to establish the existence of a solution to the continuity equation via a vanishing viscosity method, applied on the *whole* MFG system. In this way, in the second equation $D_{\mathcal{G}}u$ is replaced by $D_{\mathcal{G}}u^\sigma$ which is regular by standard regularity theory for parabolic equations and this implies the regularity of the solution of the second equation (see [15]).

Proposition 3.1. *Under assumptions* (H1) – (H4)*, problem* (3.2) *has a bounded solution m in the sense of Definition 1.1. Moreover* $m(t,\cdot) \in L^\infty([0,T], \mathcal{P}_2)$ *and* $m(t,\cdot)$ *is* $1/2$*-Hölder continuous from* $(0,T)$ *to* $\mathcal{P}_1$.

We consider the solution (u^σ, m^σ) to the following problem

$$\begin{cases} (i) \quad -\partial_t u - \sigma\Delta u + \frac{1}{2}|D_{\mathcal{G}}u|^2 = F[\overline{m}](x) & \text{in } \mathbb{R}^2 \times (0,T) \\ (ii) \quad \partial_t m - \sigma\Delta m - \mathrm{div}_{\mathcal{G}}(m D_{\mathcal{G}}u) = 0 & \text{in } \mathbb{R}^2 \times (0,T) \\ (iii) \quad m(x,0) = m_0(x), u(x,T) = G[\overline{m}(T)](x) & \text{on } \mathbb{R}^2. \end{cases} \tag{3.5}$$

Let us recall that equation (3.5)-(ii) has a standard interpretation in terms of a suitable stochastic process (see relation (3.8) below). Our aim is to find a solution to problem (3.2) letting $\sigma \to 0^+$. To this end some estimates are needed; as a first step, we establish the well-posedness of system (3.5).

Lemma 3.1. *Under assumptions* (H1) – (H4)*, for any* $\overline{m} \in C^0([0,T], \mathcal{P}_1)$*, there exists a unique bounded classical solution* (u^σ, m^σ) *to problem* (3.5)*. Moreover,* $m^\sigma > 0$.

Proof. From Lemma 3.2 here below, the solution u^σ of (3.5)-(i) is bounded in $\mathbb{R}^2 \times [0,T]$. Hence, from standard regularity results for quasilinear parabolic equations, we obtain the existence and uniqueness of a classical solution u^σ in all $\mathbb{R}^2 \times [0,T]$. Now m^σ is the classical solution of the linear equation

$$\partial_t m - \sigma\Delta m + b \cdot Dm + c_0 m = 0, \quad m(0) = m_0$$

with b and c_0 Hölder continuous coefficients. Hence, still applying classical results, we get the existence and uniqueness of a classical solution m^σ of (3.5)-(ii). From assumptions on m_0 and the classical maximum principle we get that $m^\sigma > 0$. □

Let us now prove that the functions u^σ are Lipschitz continuous and semiconcave uniformly in σ.

Lemma 3.2. *Under the same assumptions of Lemma 3.1, there exists a constant* $C > 0$*, independent of* σ *such that*

$$\|u^\sigma\|_\infty \le C, \quad \|Du^\sigma\|_\infty \le C \quad \textit{and} \quad D^2 u^\sigma \le C \qquad \forall \sigma > 0.$$

Proof. The L^∞-estimate easily follows from the Comparison Principle and assumption (H2) because the functions $w^\pm := C \pm C(T-t)$ are respectively a super- and a subsolution for (3.5)-(i) if C is sufficiently large.

We refer to [12] for the proof of the uniform Lipschitz continuity of the functions u^σ. The proof is similar to the deterministic one proved in Lemma 2.3 and it uses the representation formula by means a stochastic optimal control problem:

$$u^\sigma(x,t) = \min \mathbb{E}\left(\int_t^T \left[\frac{1}{2}|\alpha(\tau)|^2 + f(Y(\tau),\tau)\right] d\tau + g(Y(T))\right)$$

where, in $[t,T]$, $Y(\cdot)$ is governed by a stochastic differential equation

$$\begin{cases} dY_1 = \alpha_1(t)dt + \sqrt{2\sigma}\, dB_{1,t} \\ dY_2 = h(Y_1(t))\alpha_2(t)dt + \sqrt{2\sigma}\, dB_{2,t} \end{cases}, \tag{3.6}$$

where $Y(t) = x$ and B_t is a standard 2-dimensional Brownian motion. (For an analytic proof see also [27, Chapter XI]).

Let us now prove the part of the statement concerning the semiconcavity. We shall adapt the methods of [13, Lemma 5.2]. We fix a direction $v = (\alpha_1, \alpha_2)$ with $|v| = 1$ and compute the derivative of equation (3.5)-(i) twice with respect to v obtaining

$$\begin{aligned} -\partial_t \partial_{vv} u - \sigma \Delta \partial_{vv} u - \partial_{vv}(F[\overline{m}(t)](x)) &= -\partial_{vv}\left[\tfrac{1}{2}\left((\partial_1 u)^2 + h(x_1)^2(\partial_2 u)^2\right)\right] \\ &= -(D_{\mathcal{G}}\partial_v u)^2 - D_{\mathcal{G}} u \cdot D_{\mathcal{G}}\partial_{vv} u - \tfrac{1}{2}\partial_{vv}(h^2)(\partial_2 u)^2 - 4hh'\alpha_1 \partial_2 u \partial_{2v} u \\ &\le -(D_{\mathcal{G}}\partial_v u)^2 - D_{\mathcal{G}} u \cdot D_{\mathcal{G}}\partial_{vv} u + C(1 + |D_{\mathcal{G}}\partial_v u|) \end{aligned}$$

(the last inequality is due to our assumptions and to the first part of the statement). Since $-(D_{\mathcal{G}}\partial_v u)^2 + C(1 + |D_{\mathcal{G}}\partial_v u|)$ is bounded above by a constant, we deduce

$$-\partial_t \partial_{vv} u - \sigma \Delta \partial_{vv} u + D_{\mathcal{G}} u \cdot D_{\mathcal{G}}\partial_{vv} u \le C;$$

on the other hand, we have $\|\partial_{vv} u(T,\cdot)\|_\infty \le C$ by assumption (H2) and we can conclude by comparison that $\partial_{vv} u \le C'$ for a constant C' independent of σ. $\square$

Let us now prove some useful properties of the functions m^σ.

Lemma 3.3. *Under the same assumptions of Lemma 3.1, there exists a constant $K > 0$, independent of σ and of $\overline{m}$, such that:*

1. $\|m^\sigma\|_\infty \le K$,
2. $\mathbf{d}_1(m^\sigma(t_1), m^\sigma(t_2)) \le K(t_2 - t_1)^{1/2} \qquad \forall t_1, t_2 \in (0,T)$,
3. $\displaystyle\int_{\mathbb{R}^2} |x|^2\, dm^\sigma(t)(x) \le K \left(\int_{\mathbb{R}^2} |x|^2\, dm_0(x) + 1 \right) \qquad \forall t \in (0,T)$.

Proof. 1. In order to prove this L^∞ estimate, we shall argue as in [13, Appendix]; for simplicity, we drop the σ's. We note that

$$\mathrm{div}_{\mathcal{G}}(mD_{\mathcal{G}}u) = D_{\mathcal{G}}m \cdot D_{\mathcal{G}}u + m(\partial_{11}u + h^2\partial_{22}u) \le D_{\mathcal{G}}m \cdot D_{\mathcal{G}}u + Cm$$

because of the semiconcavity of u established in Lemma 3.2 yields $\partial_{ii}u \le C$ for $i = 1, 2$ (see [11, Proposition1.1.3-(e)]) and $m \ge 0$. Therefore, by assumption (H2) the function m satisfies

$$\partial_t m - \sigma\Delta m \le D_{\mathcal{G}}m \cdot D_{\mathcal{G}}u + Cm, \qquad m(x, 0) \le C;$$

using $w = Ce^{Ct}$ as supersolution (recall that C is independent of σ), we infer: $\|m\|_\infty \le w = Ce^{CT}$.

To prove Points 2 and 3 as in the proof of [12, Lemma 3.4 and 3.5], it is expedient to introduce the stochastic differential equation

$$dX_t = b(X_t, t)dt + \sqrt{2\sigma}dB_t, \qquad X_0 = Z_0 \tag{3.7}$$

where $b = (\frac{\partial u^\sigma}{\partial x_1}, h^2\frac{\partial u^\sigma}{\partial x_2})$, B_t is a standard 2-dimensional Brownian motion, and $\mathcal{L}(Z_0) = m_0$. By standard arguments, (see [23] and [22, Chapter 5])

$$m(t) := \mathcal{L}(X_t) \tag{3.8}$$

is a weak solution to (3.5)-(ii).

The rest of the proof of Points 2 and 3 follows the same arguments of [12, Lemma 3.4] and, respectively, of [12, Lemma 3.5]; therefore, we shall omit it and we refer to [12] for the detailed proof. □

Let us now prove that the u^σ's are uniformly bounded and uniformly continuous in time.

Lemma 3.4. *Under the same assumptions of Lemma 3.1, the function u^σ is uniformly continuous in time uniformly in σ.*

Proof. We shall follow the arguments in [13, Theorem 5.1 (proof)]. Let $u_f^\sigma := u^\sigma(x, T)$; by assumption (H2), there exists a constant C_1 sufficiently large such that the functions $\omega^\pm = u_f^\sigma(x) \pm C_1(T - t)$ are respectively super- and subsolution of (3.5)-(i) for any σ; actually, for $C_1 = 2C$ we have

$$-\partial_t\omega^+ - \sigma\Delta\omega^+ + \frac{1}{2}|D_{\mathcal{G}}\omega^+|^2 - F[\overline{m}](x) \ge C_1 - \sigma C - C \ge 0$$

and similarly for ω^-. Hence from the comparison principle we get

$$\|u^\sigma(x, t) - u_f^\sigma(x)\|_\infty \le C_1(T - t) \qquad \forall t \in [0, T]. \tag{3.9}$$

We look now the source term $F[\overline{m}](x)$ of (3.5)-(i). The Lipschitz continuity of F w.r.t. m (see assumption (H2)) and the uniform continuity of $\overline{m}$ imply:

$$\sup_{t\in[h,T]} \|F[\overline{m}(t)](x) - F[\overline{m}(t-h)](x)\|_\infty \le C \sup_{t\in[h,T]} \mathbf{d}_1(\overline{m}(t),\overline{m}(t-h)) =: \eta(h).$$

The function $v_h^\sigma(x,t) := u^\sigma(x,t-h) + C_1 h + \eta(h)(T-t)$ satisfies

$$\begin{aligned} -\partial_t v_h^\sigma(x,t) - \sigma\Delta v_h^\sigma(x,t) + \frac{1}{2}|D_{\mathcal{G}} v_h^\sigma(x,t)|^2 - F[\overline{m}(t)](x) + \eta(h) \\ = F[\overline{m}(t-h)](x) - F[\overline{m}(t)](x) + \eta(h) \ge 0 \qquad \forall t\in[h,T] \end{aligned}$$

and also $v_h^\sigma(x,T) = u^\sigma(x,T-h) + C_1 h \ge u^\sigma(x,T)$ by estimate (3.9); therefore, again by comparison principle, we get $u^\sigma(x,t-h) + C_1 h + \eta(h)(T-t) \ge u^\sigma(x,t)$. In a similar way we also obtain $u^\sigma(x,t-h) - C_1 h - \eta(h)(T-t) \le u^\sigma(x,t)$ accomplishing the proof. □

Proof of Proposition 3.1. We shall follow the proof of [13, Theorem 5.1] (see also [12, Theorem 4.20]). We observe that, for all $\sigma\in(0,1)$, m^σ belongs to $C^0([0,T],\mathcal{K})$ where $\mathcal{K} := \{\mu\in\mathcal{P}_1 : \mu$ satisfies Point 3 of Lemma 3.3$\}$; moreover, we recall from [12, Lemma 5.7] that $\mathcal{K}$ is relatively compact in $\mathcal{P}_1$.

Lemma 3.2 and Lemma 3.4 imply that u^σ uniformly converge to some function u and by standard stability result for viscosity solutions, the function u solves (3.3), u is Lipschitz continuous in x, $Du^\sigma \to Du$ a.e. (because of the semiconcavity estimate of Lemma 3.2 and [11, Theorem 3.3.3]), so, in particular, $D_{\mathcal{G}} u^\sigma \to D_{\mathcal{G}} u$ a.e.

By the bounds on m^σ contained respectively in Points 1 and 2 of Lemma 3.3, we obtain that, possibly passing to a subsequence, as $\sigma\to 0^+$, m^σ converge to some $m\in C^0([0,T],\mathcal{K})$ in the $C^0([0,T],\mathcal{P}_1)$ topology and in $L^\infty_{loc}((0,T)\times\mathbb{R}^2)$-weak-$*$ topology. Moreover we deduce that $m(0)=m_0$. On the other hand, since m^σ is a solution to (3.5)-(ii), for any $\psi\in C_0^\infty((0,T)\times\mathbb{R}^2)$, there holds

$$\int_0^T\int_{\mathbb{R}^2} m^\sigma\left(-\partial_t\psi - \sigma\Delta\psi + D\psi\cdot D_{\mathcal{G}} u^\sigma\right)dx\,dt = 0;$$

letting $\sigma\to 0^+$, by the L^∞_{loc}-weak-$*$ convergence of m^σ and by the convergence a.e. $D_{\mathcal{G}} u^\sigma \to D_{\mathcal{G}} u^\sigma$, we conclude that the function m solves (3.2).

Note that we proved that the solution m fulfills the estimates in Lemma 3.3. □

Remark 3.1. Note that the solution m to problem (3.2) fulfills the estimates in Lemma 3.3 with K independent of $\overline{m}$.

3.2. Uniqueness of the solution

This section is devoted to establish the following uniqueness result for problem (3.2).

Proposition 3.2. *Under assumptions* (H1) – (H4)*, problem* (3.2) *admits at most one bounded solution m in the sense of Definition 1.1. Moreover, the function m satisfies:*

$$\int_{\mathbb{R}^2} \phi\,dm(t) = \int_{\mathbb{R}^2} \phi(\overline{\gamma}_x(t))\,m_0(x)\,dx, \qquad \forall\phi\in C_0^0(\mathbb{R}^2),\ \forall t\in[0,T] \tag{3.10}$$

where, for a.e. $x \in \mathbb{R}^2$, $\overline{\gamma}_x$ *is the solution to* (1.6)*.*

In order to prove this result, it is expedient to establish some properties of the optimal trajectories for the control problem defined in Section 2 and of the value function $u(x,t)$, defined in Subsection 2.3. For any $(x,t) \in \mathbb{R}^2 \times [0,T]$, let $\mathcal{U}(x,t)$ be the set of the optimal controls of the minimization problem (OC) in Definition 2.1. We refer the reader to Appendix 6.2, for the precise definition of $\mathcal{G}$-differentiability and for its properties.

Lemma 3.5. *The following properties hold:*

1. $D_{\mathcal{G}}u(x,t)$ *exists if and only if* $\alpha(t)$ *is the same value for any* $\alpha(\cdot) \in \mathcal{U}(x,t)$*. Moreover* $D_{\mathcal{G}}u(x,t) = -\alpha(t)$ *(i.e.,* $u_{x_1}(x,t) = -\alpha_1(t)$*,* $h(x_1(t))u_{x_2}(x,t) = -\alpha_2(t)$*).*
2. *In particular, if* $\mathcal{U}(x,t)$ *is a singleton then* $D_{\mathcal{G}}u(x(s),s)$ *exists for any* $s \in [t,T]$ *where* $x(s)$ *is the optimal trajectory associated to the singleton of* $\mathcal{U}(x,t)$*.*
3. *If* x *is such that* $h(x_1) \neq 0$ *and* $D_{\mathcal{G}}u(x,t)$ *exists then there is a unique optimal trajectory starting from* x *and* $D_{\mathcal{G}}u(x,t) = -\alpha(t)$ *and hence*

$$x_1'(t) = -\partial_{x_1}u(x,t), \qquad x_2'(t) = -h^2(x_1)\partial_{x_2}u(x,t). \tag{3.11}$$

Proof. 1. We prove that if $D_{\mathcal{G}}u(x,t)$ exists then for any $\alpha(\cdot) \in \mathcal{U}(x,t)$ we have that $\alpha(t)$ is unique and $D_{\mathcal{G}}u(x,t) = -\alpha(t)$. For any $\alpha(\cdot) \in \mathcal{U}(x,t)$, let $x(\cdot)$ be the corresponding optimal trajectory. Then $x(\cdot)$ and $\alpha(\cdot)$ satisfy the necessary conditions for optimality proved in Proposition 2.1. Take $v = (v_1, v_2) \in \mathbb{R}^2$ and consider the solution $y(\cdot)$ of (1.4) with initial condition $y(t) = (x_1 + v_1, x_2 + h(x_1)v_2)$ and control α, namely

$$\begin{aligned}
y_1(s) &= x_1 + v_1 + \int_t^s \alpha_1(\tau)d\tau = x_1(s) + v_1,\\
y_2(s) &= x_2 + h(x_1)v_2 + \int_t^s h(y_1(\tau))\alpha_2(\tau)d\tau\\
&= x_2(s) + h(x_1)v_2 + \int_t^s [h(y_1(\tau)) - h(x_1(\tau))]\alpha_2(\tau)d\tau.
\end{aligned}$$

Hence there holds

$$\begin{aligned}
&u(x_1+v_1, x_2+h(x_1)v_2, t) - u(x_1,x_2,t) \le\\
&\quad\int_t^T \Bigg[f\left(x_1(s)+v_1, x_2(s)+h(x_1)v_2 + \int_t^s [h(y_1(\tau)) - h(x_1(\tau))]\alpha_2(\tau)d\tau \right)\\
&\qquad - f(x_1(s), x_2(s)) \Bigg] ds + g(y(T)) - g(x(T)).
\end{aligned}$$

For $v = t(\hat{v}_1, \hat{v}_2)$ with $|(\hat{v}_1, \hat{v}_2)| = 1$ and $t \in \mathbb{R}^+$, as $t \to 0^+$, the $\mathcal{G}$-differentiability of u at (x, t) entails

$$D_{\mathcal{G}}u(x,t) \cdot (\hat{v}_1, \hat{v}_2) \le (I_1, I_2) \cdot (\hat{v}_1, \hat{v}_2)$$

where

$$\begin{aligned}
I_1 &:= \int_t^T f_{x_1}(x(s))ds + \int_t^T \left(f_{x_2}(x(s)) \int_t^s h'(x_1(\tau))\alpha_2(\tau)d\tau \right) ds + g_{x_1}(x(T)) \\
&\quad + g_{x_2}(x(T)) \int_t^T h'(x_1(\tau))\alpha_2(\tau)d\tau \\
I_2 &:= h(x_1) \left(\int_t^T f_{x_2}(x(s))ds + g_{x_2}(x(T)) \right).
\end{aligned}$$

By the arbitrariness of $(\hat{v}_1, \hat{v}_2)$, we get

$$D_{\mathcal{G}}u(x,t) = (I_1, I_2).$$

By (2.12) and (2.6), we obtain

$$\begin{aligned}
I_1 &= \int_t^T f_{x_1}(x(s))ds + \int_t^T (p_2'(s) \int_t^s h'(x_1(\tau))\alpha_2(\tau)d\tau)ds + g_{x_1}(x(T)) \\
&\quad - p_2(x(T)) \int_t^T h'(x_1(\tau))\alpha_2(\tau)d\tau \\
&= \int_t^T f_{x_1}(x(s))ds - \int_t^T p_2(s)h'(x_1(s))\alpha_2(s)ds + g_{x_1}(x(T)) \\
&= -\alpha_1(t)
\end{aligned}$$

where the last inequality is due to (2.13) and (2.8). On the other hand, again by (2.13) and (2.8), we have

$$I_2 = -h(x_1)p_2(t) = -\alpha_2(t).$$

The last three equalities imply: $D_{\mathcal{G}}u(x,t) = -\alpha(t)$ which uniquely determines the value of $\alpha(\cdot)$ at time t.

Conversely we prove that, if for any $\alpha(\cdot) \in \mathcal{U}(x,t)$, $\alpha(t)$ is unique then $D_{\mathcal{G}}u(x,t)$ exists. To prove the $\mathcal{G}$-differentiability of $u(\cdot,t)$ in x, by the semiconcavity of u, we need to prove that $D^*_{\mathcal{G}}u(x,t)$ is a singleton (see Theorem 6.1 in Appendix 6.2 below). Let $\pi \in D^*_{\mathcal{G}}u(x,t)$. By definition of $D^*_{\mathcal{G}}u(x,t)$ there exist two sequences $\{x_n\}$, $\{\pi_n = D_{\mathcal{G}}u(x_n,t)\}$ such that

$$x_n \to x, \quad \pi_n \to \pi. \tag{3.12}$$

Consider $\alpha_n \in \mathcal{U}(x_n,t)$; by the other part of the statement (already proven), we know that

$$-\alpha_n(t) = D_{\mathcal{G}}u(x_n,t) = \pi_n. \tag{3.13}$$

Let $x_n(\cdot)$ be the trajectory associated to α_n and p_n be the corresponding costate. By Corollary 2.1-(4) and by the boundedness of h we get

$$\|x_{n1}\|_\infty + \|x_{n2}\|_\infty + \|p_{n1}\|_\infty + \|p_{n2}\|_\infty + \|\alpha_{n1}\|_\infty + \|\alpha_{n2}\|_\infty \leq C, \text{ for any } n. \tag{3.14}$$

From Corollary 2.1-(3) we can differentiate (2.8), and using (2.11)-(2.12) we get:

$$\begin{aligned}
\alpha'_{n1}(s) &= p'_{n1}(s) = -p^2_{n2}(s)h'(x_{n1}(s))h(x_{n1}(s)) + f_{x_1}(x_{n1}(s),s),\\
\alpha'_{n2}(s) &= p_{n2}(s)h'(x_{n1}(s))x'_{n1}(s) + p'_{n2}(s)h(x_{n1}(s))\\
&= p_{n2}(s)h'(x_{n1}(s))\alpha_{n1}(s) + f_{x_2}(x_{n1}(s))h(x_{n1}(s)).
\end{aligned}$$

From (3.14) we get

$$\|\alpha'_{n1}\|_\infty + \|\alpha'_{n2}\|_\infty \leq C, \text{ for any } n. \tag{3.15}$$

Hence, from Ascoli-Arzelà Theorem we have that, up to subsequences, α_n uniformly converge to some $\alpha \in C^0([t,T],\mathbb{R}^2)$. In particular, by the definition of x_{n1} and x_{n2} we get:

$$x_{n1}(s) \to x_1(s) = x_1 + \int_t^s \alpha_1(\tau)d\tau, \text{ uniformly in } [t,T],$$

$$x_{n2}(s) \to x_2(s) = x_2 + \int_t^s h(x_1(\tau))\alpha_2(\tau)d\tau \text{ uniformly in } [t,T].$$

Moreover, from stability, α is optimal, i.e. $\alpha \in \mathcal{U}(x,t)$. From the uniform convergence of the α_n we have in particular that $\alpha_n(t) \to \alpha(t)$ where $\alpha(t)$ is uniquely determined by assumption. By (3.13), we get $\pi_n \to \pi = \alpha(t)$. This implies that $D^*_{\mathcal{G}}u(x,t)$ is a singleton, then $D_{\mathcal{G}}u(x,t)$ exists and thank to the first part of the proof $D_{\mathcal{G}}u(x,t) = -\alpha(t)$.

2. If $\mathcal{U}(x,t) = \{\alpha(\cdot)\}$ then for any $s \in [t,T]$, $\alpha(s)$ is uniquely determined. Indeed, if there exists $\beta \in \mathcal{U}(x(s),s)$ the concatenation γ of α and β (see Proposition 6.1 in Appendix 6.1) is also optimal, i.e. $\gamma \in \mathcal{U}(x,t) = \{\alpha(\cdot)\}$. Then applying point 1) with $t = s$, in $x(s)$ we have that u is $\mathcal{G}$-differentiable, i.e. $D_{\mathcal{G}}u(x(s),s)$ exists.

3. From point 1), we know that for any $\alpha(\cdot) \in \mathcal{U}(x,t)$ we have that $\alpha(t)$ is unique. If we know $\alpha(t)$ and that $h(x_1(t)) = h(x_1) \neq 0$, then from (2.8) we get $p_1(t)$ and $p_2(t)$. Hence (2.9)-(2.12) is a system of differential equations with initial conditions $x_i(t)$ and $p_i(t)$, $i = 1, 2$ which admits

a unique solution $(x(s), p(s))$ where $x(s)$ is the unique optimal trajectory starting from x. Moreover still from 1) we have $D_{\mathcal{G}}u(x,t) = -\alpha(t)$ and from the dynamics (1.4) we deduce (3.11). □

Lemma 3.6. *Consider $x = (x_1, x_2) \in \mathbb{R}^2$.*

1. *Let $x(\cdot) := (x_1(\cdot), x_2(\cdot))$ be an absolutely continuous function such that*

$$x(t) = x, \tag{3.16}$$

and for almost every $s \in (t, T)$,

$$u(\cdot, s) \text{ is differentiable at } x(s), \tag{3.17}$$

and

$$\begin{aligned} x_1'(s) &= -u_{x_1}(x(s), s), \\ x_2'(s) &= -h^2(x_1(s))u_{x_2}(x(s), s), \end{aligned} \tag{3.18}$$

where u is the solution of (2.16). Then the control law $\alpha(s) = (\alpha_1(s), \alpha_2(s))$, with

$$\alpha_1(s) = -u_{x_1}(x(s), s), \ \ \alpha_2(s) = -h(x_1(s))u_{x_2}(x(s), s)$$

is optimal for $u(x, t)$.

2. *If $u(\cdot, t)$ is $\mathcal{G}$-differentiable at x and $h(x_1) \neq 0$ then problem (1.6) has a unique solution corresponding to the optimal trajectory.*

Proof. We shall adapt the arguments of [12, Lemma 4.11]. Fix $(t, x) \in (0, T) \times \mathbb{R}^2$ and consider an absolutely continuous solution $x(\cdot)$ to (3.18); note that this implies that Du exists at $(x(s), s)$ for a.e. $s \in (t, T)$. Since u is Lipschitz continuous (see Lemma 2.3) and h is bounded, also the function $x(\cdot)$ is Lipschitz continuous and, consequently, also $u(x(\cdot), \cdot)$ is Lipschitz. For a.e. $s \in (t, T)$ there hold: $i)$ $Du(x(s), s)$ exists, $ii)$ equation (3.18) holds, $iii)$ the function $u(x(\cdot), \cdot)$ admits a derivative at s. Fix such a s.

The Lebourg Theorem for Lipschitz function (see [17, Thm 2.3.7] and [17, Thm 2.5.1]) ensures that, for any $h \in \mathbb{R}$ small, there exists (y_h, s_h) in the segment $((x(s), s), (x(s+h), s+h))$ and $(\xi_x^h, \xi_t^h) \in coD^*_{x,t}u(y_h, s_h)$ such that

$$u(x(s+h), s+h) - u(x(s), s) = \xi_x^h \cdot (x(s+h) - x(s)) + \xi_t^h h \tag{3.19}$$

(here, "co" stands for the convex hull and $D^*_{x,t}u(y_h, s_h)$ is the Euclidean reachable gradient both in x and in t, see [4, eq. (4.4)]). The Caratheodory theorem (see [11, Thm A.1.6]) guarantees that there exist $(\lambda^{h,i}, \xi_x^{h,i}, \xi_t^{h,i})_{i=1,\dots,4}$ such that $\lambda^{h,i} \geq 0$, $\sum_{i=1}^4 \lambda^{h,i} = 1$, $(\xi_x^{h,i}, \xi_t^{h,i}) \in D^*_{x,t}u(y_h, s_h)$ and $(\xi_x^h, \xi_t^h) = \sum_{i=1}^4 \lambda^{h,i}(\xi_x^{h,i}, \xi_t^{h,i})$. We claim that $\xi_x^{h,i} \to D_x u(x(s), s)$ as $h \to 0$. Actually, let l_x^i be any cluster point of $\{\xi_x^{h,i}\}$. By a diagonalization process, there exist (x_n, t_n) such that $(x_n, t_n) \to (x(s), s)$ and $D_x u(x_n, t_n)$ exist and converge to l_x^i as $n \to \infty$. By [12, Lemma 4.6] and [11, Proposition 3.1.5-(c)], we have $l_x^i = \lim_n D_x u(x_n, t_n) \in D_x^+ u(x(s), s) = D_x u(x(s), s)$; our claim is proved.

On the other hand, since u is a viscosity solution to equation (2.16), by [4, Proposition II.1.9], we obtain

$$-\xi_t^{h,i}+\frac{1}{2}(\xi_{x,1}^{h,i})^2+\frac{1}{2}h(y_{h,1})^2(\xi_{x,2}^{h,i})^2=f(y_h,s_h);$$

in particular, as $h\to 0$, we deduce

$$\xi_t^h=\frac{1}{2}\sum_{i=1}^{4}\lambda^{h,i}(\xi_{x,1}^{h,i})^2+\frac{1}{2}h(y_{h,1})^2\sum_{i=1}^{4}\lambda^{h,i}(\xi_{x,2}^{h,i})^2-f(y_h,s_h)\to\frac{1}{2}|D_{\mathcal{G}}u(x(s),s)|^2-f(x(s),s). \tag{3.20}$$

Dividing (3.19) by h and letting $h\to 0$, by equations (1.6) and (3.20), we infer

$$\begin{aligned}\frac{d}{ds}u(x(s),s)&=D_xu(x(s),s)\cdot x'(s)+\frac{1}{2}|D_{\mathcal{G}}u(x(s),s)|^2-f(x(s),s)\\&=-\frac{1}{2}|D_{\mathcal{G}}u(x(s),s)|^2-f(x(s),s)=\frac{1}{2}|\alpha|^2-f(x(s),s)\qquad\text{a.e. } s\in(t,T)\end{aligned}$$

(recall: $-\alpha(s)=D_{\mathcal{G}}u(x(s),s)$). Integrating this equality on $[t,T]$ and taking into account the final datum of (2.16), we obtain

$$u(x,t)=\int_t^T\frac{1}{2}|\alpha|^2+f(x(s),s)ds+g(x(T)).$$

Observe that $x(\cdot)$ satisfies the dynamics (1.4) with our choice of $\alpha(s)$; therefore, the last equality implies that $x(\cdot)$ is an optimal trajectory with optimal control $\alpha(s)=-D_{\mathcal{G}}u(x(s),s)$.

Let us now prove the last part of the statement. By Point 3 of Lemma 3.5, there exists a unique optimal trajectory $x(\cdot)$ starting from x at time t; moreover, by Corollary 2.2, for any $s\in(t,T]$ there exists a unique optimal trajectory starting from $x(s)$ which is the restriction of $x(\cdot)$ to $[s,T]$. Then, from the representation of the optimal controls (2.14), there exists a unique optimal control $\alpha(\cdot)$ and, from points 1 and 2 of Lemma 3.5, $D_{\mathcal{G}}u(x(s),s)$ exists and $D_{\mathcal{G}}u(x(s),s)=-\alpha(s)$, i.e. $x(\cdot)$ is a solution of (1.6). Moreover this $x(\cdot)$ is the unique solution still because of Point 3 of Lemma 3.5. □

Proof of Proposition 3.2. We shall argue following the techniques of [15, Proposition A.1] which rely on the superposition principle and on the disintegration of a measure (see [3]). We denote by Γ_T the set of continuous curve $C^0([0,T],\mathbb{R}^2)$ and, for any $t\in[0,T]$, we introduce the evaluation map: $e_t:\Gamma_T\to\mathbb{R}^2$ as $e_t(\gamma):=\gamma(t)$. When we say "for a.e." without specifying the measure, we intend w.r.t. the Lebesgue measure.

Let $m\in C^0([0,T],\mathcal{P}_1(\mathbb{R}^2))$ be a solution of problem (3.2) in the sense of distributions; in other words, it is a solution to the continuity equation (3.1). We observe that assumption [3, eq.(8.1.20)] is fulfilled because both Du and h are bounded and $m_t:=m(t,\cdot)$ is a measure (see [3, pag.169]); hence we can invoke the superposition principle (see [3, Theorem 8.2.1] and also [3, pag. 182]). This principle and the disintegration theorem (see [3, Theorem 5.3.1]) entail that there exist probability measures η and $\{\eta_x\}_{x\in\mathbb{R}^2}$ on Γ_T such that

$i)$ $e_t\#\eta = m_t$ and, in particular, $e_0\#\eta = m_0$

$ii)$ $\eta_x(\{\gamma \in \Gamma_T : \gamma$ solves (1.6) with $t = 0$ and $x = (x_1, x_2)\}) = 1$ for m_0-a.e. x

$iii)$ $\eta = \int\limits_{\mathbb{R}^2} \eta_x \, dm_0(x)$.

We recall from assumption (H4) that m_0 is absolutely continuous; hence, by assumption (H2) and meas$\{x \in \mathbb{R}^2 : h(x_1) = 0\} = 0$, the optimal synthesis in Lemma 3.6 ensures that for a.e. $x \in \mathbb{R}^2$ the solution $\overline{\gamma}_x$ to (1.6) with $t = 0$ and $x = (x_1, x_2)$ is unique and exists because it is the optimal trajectory for the control problem. Therefore, for a.e. $x \in \mathbb{R}^2$, η_x coincides with $\delta_{\overline{\gamma}_x}$. In conclusion, for any function $\phi \in C_0^0(\mathbb{R}^2)$, we have

$$\int\limits_{\mathbb{R}^2} \phi \, dm_t = \int\limits_{\Gamma_T} \phi(e_t(\gamma)) d\eta(\gamma) = \int\limits_{\mathbb{R}^2} \left(\int\limits_{e_0^{-1}(x)} \phi(e_t(\gamma)) d\eta_x(\gamma) \right) dm_0(x)$$
$$= \int\limits_{\mathbb{R}^2} \phi(\overline{\gamma}_x(t)) m_0(x) \, dx.$$

Since the integrand in the last term is uniquely defined up to a set of null measure, also the first term is uniquely defined; consequently, m is uniquely defined. □

In the following corollary we use the previous characterization to prove the Lipschitz regularity of m.

Corollary 3.1. *The unique bounded solution m to problem* (3.2) *is a Lipschitz continuous map from* $[0, T]$ *to* $\mathcal{P}_1(\mathbb{R}^2)$ *with a Lipschitz constant bounded by* $\|Du\|_\infty \|h^2\|_\infty$.

Proof. Let m be the unique solution to problem (3.2) as in Proposition 3.1 and Proposition 3.2. Fix ϕ, a 1-Lipschitz continuous function on $\mathbb{R}^2$. By relation (3.10), for any $t_1, t_2 \in [0, T]$, we infer

$$\int\limits_{\mathbb{R}^2} \phi \, dm_{t_1} - \int\limits_{\mathbb{R}^2} \phi \, dm_{t_2} = \int\limits_{\mathbb{R}^2} \phi(\overline{\gamma}_x(t_1)) - \phi(\overline{\gamma}_x(t_2)) m_0(x) \, dx$$
$$\leq \int\limits_{\mathbb{R}^2} \left|\overline{\gamma}_x(t_1) - \overline{\gamma}_x(t_2)\right| m_0(x) \, dx$$
$$\leq \|Du\|_\infty \|h^2\|_\infty |t_1 - t_2|$$

where the last relation is due to the definition of $\overline{\gamma}$ as solution to problem (1.6) and to the boundedness of Du and of h. Hence, passing to the $\sup_\phi$ in the previous inequality, the Kantorovich-Rubinstein theorem (see, for instance, [12, Theorem 5.5]) ensures

$$\mathbf{d}_1(m_{t_1}, m_{t_2}) \leq \|Du\|_\infty \|h^2\|_\infty |t_1 - t_2|. \quad \square$$

Proof of Theorem 3.1. The existence of m follows from Proposition 3.1, the uniqueness and the representation formula comes from Proposition 3.2 and the Lipschitz regularity is proved in Corollary 3.1 here above. □

4. Proof of the main Theorem

This section is devoted to the proof of our main Theorem 1.1.

Proof of Theorem 1.1. 1. We shall argue following the proof of [12, Theorem 4.1]. Consider the set

$$\mathcal{C} := \{m \in C^0([0,T],\mathcal{P}_1);\ m(0) = m_0\}$$

endowed with the norm of $C^0([0,T];\mathcal{P}_1)$. Observe that it is a nonempty closed and convex subset of $C^0([0,T];\mathcal{P}_1)$. We introduce a map $\mathcal{T}$ as follows: to any $m \in \mathcal{C}$ we associate the solution u to problem (2.16) with $f(x,t) = F[m(t)](x)$ and $g(x) = G[m(T)](x)$ and to this u we associate the solution $\mu =: \mathcal{T}(m)$ to problem (3.2). By Theorem 3.1 the function $\mathcal{T}(m)$ belongs to $\mathcal{C}$ hence $\mathcal{T}$ maps $\mathcal{C}$ into itself. We claim that the map $\mathcal{T}$ has the following properties:

(a) $\mathcal{T}$ is a continuous map with respect to the norm of $C^0([0,T];\mathcal{P}_1)$
(b) $\mathcal{T}$ is a compact map.

(a) It suffices to follow the same arguments as those in [12, Lemma 4.19]) or in [2, Theorem 2.1]).

(b) Since $\mathcal{C}$ is closed, it is enough to prove that $\mathcal{T}(\mathcal{C})$ is a precompact subset of $C^0([0,T];\mathcal{P}_1)$. Let $(\mu_n)_n$ be a sequence in $\mathcal{T}(\mathcal{C})$ with $\mu_n = \mathcal{T}(m_n)$ for some $m_n \in \mathcal{C}$; we wish to prove that, possibly for a subsequence, μ_n converges to some μ in the $C^0([0,T];\mathcal{P}_1(\mathbb{R}^2))$-topology as $n \to \infty$. By Remark 3.1, the functions $\mathcal{T}(m_n)$ satisfy the estimates in Lemma 3.3 with a constant independent of n. Since the subsets of $\mathcal{P}_1$ whose elements have uniformly bounded second moment are relatively compact in $\mathcal{P}_1$ (see [12, Lemma 5.7]), Remark 3.1 and Proposition 3.1 ensures that the sequence $(\mathcal{T}(m_n))_n$ is uniformly bounded in $C^{1/2}([0,T];\mathcal{P}_1)$ and $L^\infty(0,T;\mathcal{P}_2)$. By arguing as in the proof of Proposition 3.1, we obtain that, possibly for a subsequence (still denoted by $\mathcal{T}(m_n)$), $\mathcal{T}(m_n)$ converges to some μ in the $C^0([0,T];\mathcal{P}_1(\mathbb{R}^2))$-topology (for more details see also [2, Theorem 2.1]). Invoking Schauder fixed point Theorem, we accomplish the proof of 1.

2. Theorem 3.1 ensures that, if (u,m) is a solution of (1.1), for any function $\phi \in C^0_0(\mathbb{R}^2)$, we have

$$\int_{\mathbb{R}^2} \phi \, dm(t) = \int_{\mathbb{R}^2} \phi(\overline{\gamma}_x(t)) m_0(x)\, dx \tag{4.1}$$

where $\overline{\gamma}_x$ is the solution of (1.6) (with $t = 0$ and $x = (x_1, x_2)$) and it is uniquely defined for a.e. $x \in \mathbb{R}^2$. □

Remark 4.1. As in [12, Theorem 4.20] also the vanishing viscosity method may be applied to prove the existence of a solution of system (1.1). Actually, it suffices to follow the same arguments of Section 3.1 with $F[\overline{m}](x)$ and $G[\overline{m}(T)](x)$ replaced respectively by $F[m^\sigma](x)$ and $G[m^\sigma(T)](x)$. Note also that Lemma 3.3 ensures that the function $m^\sigma \in C^0([0,T],\mathcal{P}_1)$.

Because of the degenerate term h, we cannot directly deduce the representation formula (4.1) invoking the results in [12], but we can apply the results of Section 3.2.

5. The d-dimensional case

In this section we show that the results proved before can be generalized to the d-dimensional case. We consider dynamics governed by a triangular matrix whose coefficients have a suitable structure (see (5.3) below). This type of matrix allows us to obtain a L^∞ estimate of the optimal control associated to the Hamilton-Jacobi equation (5.1)-(i) which plays a crucial role in our argument. Moreover the assumption (H3′) here below on the set $\mathcal{Z}$ allows us to apply the representation formula also in this case. The proofs rely on the same arguments of the problem studied in the previous sections; hence we only emphasize the main differences.

5.1. Assumptions and main result

We consider the following Mean Field Game system

$$\begin{cases} (i) & -\partial_t u + H(x, Du) = F[m(t)](x) & \text{in } \mathbb{R}^d \times (0,T) \\ (ii) & \partial_t m - \operatorname{div}(m\, \partial_p H(x, Du)) = 0 & \text{in } \mathbb{R}^d \times (0,T) \\ (iii) & m(x,0) = m_0(x), u(x,T) = G[m(T)](x) & \text{on } \mathbb{R}^d, \end{cases} \tag{5.1}$$

where, for $x = (x_1, \cdots, x_d)$ and $p = (p_1, \cdots, p_d)$, the function $H(x, p)$ is

$$H(x,p) = \frac{1}{2}|pB(x)|^2 \tag{5.2}$$

with $B(x) = B(x_1, ..., x_d)$ is a $d \times d$ matrix of the form

$$\begin{pmatrix} h_{11} & 0 & 0 & 0 & \cdots & 0 \\ h_{21}(x_1) & h_{22}(x_1) & 0 & 0 & \cdots & 0 \\ h_{31}(x_1) & h_{32}(x_1,x_2) & h_{33}(x_1,x_2) & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ h_{d1}(x_1) & h_{d2}(x_1,x_2) & h_{d3}(x_1,x_2,x_3) & \cdots & h_{d(d-1)}(x_1,\dots,x_{d-1}) & h_{dd}(x_1,\dots,x_{d-1}) \end{pmatrix} \tag{5.3}$$

namely the matrix $B(x)$ is triangular inferior, the first column has terms which depend only on x_1 except h_{11} which is constant; for $j \geq 2$ the diagonal terms h_{jj} depend on the variables $(x_1, x_2, \dots, x_{j-1})$ and the terms h_{ij} with $i > j$ depend on the variables $(x_1, x_2, \dots, x_j)$. More precisely

$$h_{ij} = \begin{cases} 0 & \text{if } i < j \\ h_{jj}(x_1, x_2, \dots, x_{j-1}), & \text{if } i = j \\ h_{ij}(x_1, x_2, \dots, x_j), & \text{if } i > j. \end{cases}$$

We introduce the determinant function $\Delta(x)$

$$\Delta(x) = \det B(x) = h_{11}h_{22}(x_1)\cdots h_{dd}(x_1, \dots, x_{d-1}). \tag{5.4}$$

We shall assume the following hypotheses.

(H1′) The functions F and G are real-valued function, continuous on $\mathcal{P}_1(\mathbb{R}^d) \times \mathbb{R}^d$;

(H2′) The map $m \to F[m](\cdot)$ is Lipschitz continuous from $\mathcal{P}_1(\mathbb{R}^d)$ to $C^2(\mathbb{R}^d)$; moreover, there exists $C \in \mathbb{R}$ such that

$$\|F[m](\cdot)\|_{C^2}, \|G[m](\cdot)\|_{C^2} \le C, \qquad \forall m \in \mathcal{P}_1(\mathbb{R}^d),$$

(H3′) h_{11} is a constant, the functions $h_{ij} : \mathbb{R}^{i-1} \to \mathbb{R}$ are $C^2(\mathbb{R}^{i-1})$ with $\|h_{ij}\|_{C^2} \le C$ and $\mathcal{Z} := \{z \in \mathbb{R}^d \ : \ \Delta(z) = 0\}$ has null measure (hence $h_{11} \ne 0$).

(H4′) m_0 has a compactly supported density $m_0 \in C^{2,\delta}(\mathbb{R}^d)$, for a $\delta \in (0, 1)$.

The MFG system (5.1) arises when the generic player chooses the control $\alpha \in L^2([t, T]; \mathbb{R}^d)$ in order to minimize the cost

$$\int_t^T \left[\frac{1}{2}|\alpha(\tau)|^2 + F[m(\tau)](x(\tau))\right] d\tau + G[m(T)](x(T)) \tag{5.5}$$

where, in $[t, T]$, its dynamics $x(\cdot) \in \mathbb{R}^d$ are governed by

$$\begin{cases} x'(s) = \alpha(s)B^T(x(s)), & a.e.\ s \in (t, T), \\ x(t) = x \end{cases} \tag{5.6}$$

with $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{R}^d$.

Note that, from (5.2), we have $\partial_p H(x, p) = p\, B(x)\, B^T(x)$. The main theorem is

Theorem 5.1. *Under the above assumptions:*

1. *System* (5.1) *has a solution* (u, m),
2. *m is the push-forward of m_0 through the characteristic flow*

$$x'(s) = -Du(x(s), s)B(x(s))B^T(x(s)), \qquad x(0) = x. \tag{5.7}$$

To prove Theorem 5.1 we follow the steps used in the previous sections.

5.2. *Optimal control problem*

For every $0 \le t \le T$ and $x \in \mathbb{R}^d$ we consider the following optimal control problem

Definition 5.1 *(Optimal Control Problem (OCd)).* Minimize $J_t(x(\cdot), \alpha)$ as in (2.1) subject to $(x(\cdot), \alpha(\cdot)) \in \mathcal{A}(x, t)$, where

$$\mathcal{A}(x, t) := \left\{(x(\cdot), \alpha(\cdot)) \in AC([t, T]; \mathbb{R}^d) \times L^2([t, T]; \mathbb{R}^d) : (5.6) \text{ holds}\right\}. \tag{5.8}$$

We assume that the functions f, g and the matrix $B = (h_{ij})$ satisfy the following assumptions.

Hypothesis 5.1. $f \in C^0([0,T], C^2(\mathbb{R}^d))$ and there exists a constant C such that

$$\|f(\cdot,t)\|_{C^2(\mathbb{R}^d)} + \|g\|_{C^2(\mathbb{R}^d)} + \|h_{ij}\|_{C^2(\mathbb{R}^{i-1})} \leq C, \qquad \forall t \in [0,T].$$

The set $\mathcal{Z} := \{z \in \mathbb{R}^d \ : \ \Delta(z) = 0\}$ has null measure.

Note that also in this case Remark 2.2 still holds. The definition of the value function $u(x,t)$ is the same as in the 2-dim case (2.3).

The application of the Maximum Principle yields the following necessary conditions.

Proposition 5.1 *(Necessary conditions for optimality). Let* (x^*, α^*) *be optimal for* (OCd). *There exists an arc* $p \in AC([t,T]; \mathbb{R}^d)$*, hereafter called the costate, such that*

1. *The pair* (α^*, p) *satisfies the adjoint equations*

$$p'(s) = -D_x(p \cdot \alpha^*(s) B^T(x^*(s))) + D_x f(x^*(s), s) \ \ a.e.\ s \in [t,T], \tag{5.9}$$

the transversality condition

$$-p(T) = Dg(x^*(T)) \tag{5.10}$$

together with the maximum condition

$$\max_{\alpha=(\alpha_1,\ldots,\alpha_d)\in\mathbb{R}^d} (p(s) \cdot \alpha B^T(x^*(s))) - \frac{|\alpha|^2}{2} =$$
$$= (p(s) \cdot \alpha^*(s) B^T(x^*(s))) - \frac{|\alpha^*(s)|^2}{2} \ \ a.e.\ s \in [t,T]. \tag{5.11}$$

2. *The optimal control* α^* *is given by*

$$\alpha^*(s) = p(s)\, B(x^*(s)), \ \ a.e\ on\ [t,T]. \tag{5.12}$$

3. *The pair* (x^*, p) *satisfies the system of differential equations for a.e.* $s \in [t,T]$

$$\begin{cases} (i) \quad x' = p\, B(x) B^T(x), \\ (ii) \quad p' = -\dfrac{D_x |p\, B(x)|^2}{2} + D_x f(x,s) \end{cases} \tag{5.13}$$

with the mixed boundary conditions

$$x^*(t) = x, \quad p(T) = -Dg(x^*(T)). \tag{5.14}$$

Corollary 5.1 *(Feedback control and regularity). Let Hypothesis 5.1 hold. Let* (x^*, α^*) *be optimal for (OCd), and* p *be the corresponding costate as in Proposition 5.1. Then:*

1. *The unique solution of the Cauchy problem*

$$p'(s) = -\left(\frac{D_x|p\,B(x^*(s))|^2}{2}\right) + D_x f(x^*(s), s), \quad p(T) = -Dg(x^*(T))$$

   is the costate p associated to (x^*, α^*).
2. *The optimal α^* is a feedback control and it is uniquely expressed by $\alpha^* = p\,B(x^*)$.*
3. *The functions x^* and α^* are of class C^1. In particular equations (5.9) and (5.13) hold for every $s \in [t, T]$.*
4. *There is a constant C independent of (x, t) such that $\|\alpha^*\|_\infty + \|p\|_\infty \le C$.*
5. *Assume that, for some $k \in \mathbb{N}$, $h_{ij} \in C^{k+1}$ for all $d \ge i \ge j \ge 1$ and that $D_x f \in C^k$. Then p and α^* are of class C^{k+1} and x^* is of class C^{k+2}.*

Proof. The proof follows the lines as in Corollary 2.1 for $d = 2$. We only outline the proof of point 4 where the method is slightly different. The structure of the matrix B, given in (5.3), allows us to obtain a diagonal system of differential equations in (5.13)-(ii). More precisely since the functions h_{ij} in the matrix B do not depend on the last variable x_d, the last coordinate of $D_x|p(s)B(x)|^2$ is 0, then the last component of (5.13)-(ii) is $p'_d(s) = f_{x_d}(x^*(s), s)$; this equation and the hypothesis on the data imply that p_d is bounded uniformly on (x, t). Using again the structure of that matrix B, remarking that the $(d-1)$-th component of $D_x|p(s)B(x)|^2$ is

$$\frac{\partial}{\partial x_{d-1}}(p_{d-1}\,h_{(d-1)(d-1)}(x_1, \dots, x_{d-2}) + p_d\,h_{d(d-1)}(x_1, \dots, x_{d-1}))^2.$$

Hence still from (5.13)-(ii) we get

$$p'_{d-1} = f_{x_{d-1}}(x^*(s), s) - (p_{d-1}\,h_{(d-1)(d-1)} + p_d\,h_{d(d-1)})\,p_d\,\frac{\partial h_{d(d-1)}}{\partial x_{d-1}}.$$

Now using the uniform boundedness of p_d obtained just before and the assumptions on h_{ij}, f and g, the differential equation satisfied by p_{d-1} is linear with bounded coefficients. Hence we obtain the uniform boundedness of p_{d-1}. Applying the same procedure iteratively, we get that every component p_i of p is uniformly bounded. Once obtained p we immediately get the uniform bound on optimal α^* thanks to (5.12). □

Theorem 5.2 *(Uniqueness of the optimal trajectory after the initial time). Under Hypothesis 5.1, let x^* be an optimal trajectory for $u(x, t)$.*

1. *Assume that $\Delta(x^*(\tau)) \neq 0$ for some $t < \tau < T$. For every $\tau \le r < T$ there are no other optimal trajectories for $u(x^*(r), r)$, other than x^*, restricted to $[r, T]$.*
2. *Assume that $\Delta(x) = 0$. Let t_{x^*} be defined by*

$$t_{x^*} := \sup\{r \in [t, T] : \Delta(x^*) = 0 \text{ on } [t, r]\}.$$

   For every $r > t_{x^*}$ there are no optimal trajectories starting from $x^*(r)$ at time r, other than x^* restricted to $[r, T]$.

Proof. 1. For $r \in [\tau, T[$, let y^* be an optimal trajectory starting from $x^*(r)$ at time r. The concatenation z^* of x^* with y^* at r is optimal for $u(x,t)$. Let p and q be the costates associated to x^* and, respectively, to z^*. Both (x^*, p) and (z^*, q) are solutions to the same Cauchy problem (5.13). Corollary 5.1 shows that x^* and z^* are of class C^1. Since $\Delta(x^*(\tau)) \neq 0$, then the matrix $B(x^*(\tau))B^T(x^*(\tau))$ is invertible whose inverse we denote $\beta(\tau) := [B(x^*(\tau))B^T(x^*(\tau))]^{-1}$. Hence, from (5.13), it is possible to write

$$p(\tau) = \beta(\tau)(x^*(\tau))'.$$

Since $x^* = z^*$ on $[t, \tau]$, the fact that $\tau > t$ and the C^1 regularity of x^* and z^* imply

$$p(\tau) = \beta(\tau)(x^*(\tau))' = \lim_{s\to\tau^-} \beta(s)(x^*(s))' = \lim_{s\to\tau^-} \beta(s)(z^*(s))' = \beta(\tau)(z^*(\tau))' = q(\tau).$$

Therefore, both (x^*, p) and (z^*, q) are C^1 solutions to the same Cauchy problem on $[t, T]$, with Cauchy data at τ, for the first order differential system (5.13). The regularity assumptions on f, h and Cauchy Lipschitz Theorem guarantee the uniqueness of the solution. Thus $x^* = z^*$ on $[\tau, T]$, from which we obtain the desired equality $x^* = y^*$ on $[r, T]$.

2. We assume $t_{x^*} < T$, otherwise the claim is trivial. We deduce that there exists $\tau \in [t_{x^*}, r]$ satisfying $\Delta(\tau) \neq 0$. Point 1 of Theorem 5.2 yields the conclusion. □

Corollary 5.2. *Let x^* be an optimal trajectory for (OCd). If $\Delta(x) \neq 0$, for every $0 < r < T$ there are no other optimal trajectories starting from $x^*(r)$ at time r, other than x^*, restricted to $[r, T]$.*

5.3. *Proof of the Theorem 5.1*

Following the same procedure as in Section 2.3, taking account of Corollary 2.1, we can prove that the solution u of the Hamilton-Jacobi equation (5.1)-(i) with m fixed is bounded in $\mathbb{R}^d \times [0, T]$, Lipschitz continuous with respect to x and t and it is semiconcave with respect to x.

Moreover following the arguments used in Lemma 3.5 and Lemma 3.6 and the results on $\mathcal{G}$-differentiability stated in subsection 6.3 of the Appendix, we get the optimal synthesis:

Proposition 5.2 *(Optimal synthesis). Consider $x \in \mathbb{R}^d$. Let $x(\cdot)$ be an absolutely continuous function such that $x(t) = x$, and for almost every $s \in (t, T)$, $u(\cdot, s)$ is differentiable at $x(s)$, and*

$$x'(s) = -Du(x(s), s)\, B(x(s))\, B^T(x(s)),$$

where u is the solution of (5.16). Under these assumptions, the control law

$$\alpha(s) = -Du(x(s), s)B(x(s))$$

is optimal for $u(x,t)$. If $u(\cdot, t)$ is $\mathcal{G}$-differentiable at x and $\Delta(x) \neq 0$ then problem (5.7) has a unique solution corresponding to the optimal trajectory.

The next step is to study the problem

$$\begin{cases} \partial_t m - \operatorname{div}(m\, Du\, B(x)\, B^T(x)) = 0, & \text{in } \mathbb{R}^d \times (0,T), \\ m(x,0) = m_0(x), & \text{on } \mathbb{R}^d, \end{cases} \tag{5.15}$$

where u is a solution to problem

$$\begin{cases} -\partial_t u + \frac{1}{2}|Du\, B(x)|^2 = F[\overline{m}(t)](x) & \text{in } \mathbb{R}^d \times (0,T), \\ u(x,T) = G[\overline{m}(T)](x), & \text{on } \mathbb{R}^d, \end{cases} \tag{5.16}$$

where the function $\overline{m}$ is fixed and is in $C^0([0,T], \mathcal{P}_1)$.

Using the same arguments of Section 3, we prove the existence and uniqueness of the solution m and by the superposition principle we get the representation formula of m as the push-forward of some measure on $C^0([0,T], \mathbb{R}^d)$ through the flow defined in (5.7).
To adapt the proof of Lemma 3.3-(1), it is important to point out that, thanks to the boundedness of the coefficients of B and to the semiconcavity and Lipschitz continuity of u, there holds

$$\operatorname{div}(DuBB^T) = \sum_k \partial_k \left(\sum_{ij} \partial_i u\, h_{ij}\, h_{kj} \right) = \left[\sum_j h_j\, D^2 u\, h_j^T + \sum_{ijk} \partial_i u\, \partial_k (h_{ij}\, h_{kj}) \right] \leq C,$$

where $C > 0$, $h_j = (h_{ij})_i$. Note that we can repeat the arguments as in Proposition 3.2 since we have the optimal synthesis and the assumption meas$\mathcal{Z} = 0$.

With all these ingredients we are able to follow the arguments of Section 4 to infer the proof of the main Theorem 5.1.

6. Appendix

6.1. Concatenation of optimal trajectories and the Dynamic Programming Principle

We introduce the notion of concatenation of trajectories and prove a variant of the Dynamic Programming Principle.

Definition 6.1. For $0 \leq t \leq r < T$, let $\varphi : [t,T] \to \mathbb{R}^n$ and $\psi : [r,T] \to \mathbb{R}^n$. The concatenation of φ with ψ at r is the function $\xi : [t,T] \to \mathbb{R}^n$ defined by

$$\xi = \varphi \text{ on } [t,r], \qquad \xi = \psi \text{ on } [r,T].$$

Proposition 6.1 *(Dynamic Programming Principle). Let x^* be optimal for $u(x,t)$, and $r \in [t,T]$. Let α^* be optimal control for x^*.*

1. *Let y^* be optimal for $u(x^*(r), r)$. The concatenation of x^* with y^* at r is optimal for $u(x,t)$ and, moreover,*

$$u(x,t) = u(x^*(r), r) + \int_t^r \frac{1}{2}|\alpha^*(s)|^2 + f(x^*(s), s)\, ds; \tag{6.1}$$

2. *The trajectory x^*, restricted to $[r, T]$, is optimal for $u(x^*(r), r)$;*
3. *The couple (x^*, α^*), restricted to $[t, r]$, is optimal for the following optimal control problem with prescribed endpoints:*

$$\textit{Minimize } I_{t,r}(x, \alpha) := \int_t^r \frac{1}{2}|\alpha(s)|^2 + f(x(s), s)\, ds,$$

with $(x(\cdot), \alpha)$ subject to (1.4) and $x(r) = x^*(r)$.
4. *The Dynamic Programming Principle holds:*

$$u(x,t) = \min_{(x(\cdot),\alpha)\in\mathcal{A}(x,t)} \left\{ u(x(r), r) + \int_t^r \frac{1}{2}|\alpha(s)|^2 + f(x(s), s)\, ds \right\}. \tag{6.2}$$

Proof. 1. Let β^* be optimal control for y^*. Let (z^*, γ^*) be the concatenation of (x^*, α^*) with (y^*, β^*) at r: clearly (z^*, γ^*) is admissible for (OC) of Definition 2.1. The minimality of (x^*, α^*) for $u(x, t)$, and that of (y^*, β^*) for $u(x^*(r), r)$, directly yield

$$\begin{aligned}
u(x,t) &= \int_t^r \frac{1}{2}|\alpha^*|^2 + f(x^*, s)ds + \left(\int_r^T \frac{1}{2}|\alpha^*|^2 + f(x^*, s)ds + g(x^*(T)) \right) \\
&\geq \int_t^r \frac{1}{2}|\alpha^*|^2 + f(x^*, s)ds + u(x^*(r), r) \\
&= \int_t^r \frac{1}{2}|\alpha^*|^2 + f(x^*, s)ds + \left(\int_r^T \frac{1}{2}|\beta^*|^2 + f(y^*, s)ds + g(y^*(T)) \right) \\
&= J_t(z^*, \gamma^*) \geq u(x,t),
\end{aligned}$$

so that the above inequalities are actually equalities, proving (6.1) and the optimality of (z^*, γ^*).

2. Let (y, β) be admissible for $u(x^*(r), r)$. Let (z, γ) be the concatenation of (x^*, α^*) with (y, β) at r. The conclusion follows from the following inequality:

$$0 \leq J_t(z, \gamma) - J_t(x^*, \alpha^*) = J_r(y, \beta) - J_r(x^*, \alpha^*).$$

3. Assume that $(x(\cdot), \alpha)$ is admissible for $u(x, t)$, in the interval $[t, r]$, i.e., satisfies (1.4) together with the *endpoint* condition $x(r) = x$. Then the concatenation (z, γ) of $(x(\cdot), \alpha)$ with (x^*, α^*), restricted to $[r, T]$, at r is admissible. The minimality of (x^*, α^*) implies that

$$J_t(x^*, \alpha^*) \leq J_t(z, \gamma). \tag{6.3}$$

Now

$$J_t(x^*, \alpha^*) = I_{t,r}(x^*, \alpha^*) + J_r(x^*, \alpha^*), \quad J_t(z, \gamma) = I_{t,r}(x(\cdot), \alpha) + J_r(x^*, \alpha^*).$$

It follows from (6.3) that $I_{t,r}(x^*, \alpha^*) \leq I_{t,r}(x(\cdot), \alpha)$.

4. We shall follow arguments similar to those of [4, Proposition III.2.5]. Let $(x(\cdot), \alpha)$ be admissible and (y^*, β^*) be optimal for $u(x(r), r)$. Let (z, γ) be the concatenation of $(x(\cdot), \alpha)$ with (y^*, β^*) at r. Since (z, γ) is admissible we get

$$u(x,t) \leq J_t(z, \gamma) = \int_t^r \frac{1}{2}|\alpha(s)|^2 + f(x(s), s)\, ds + u(x(r), r),$$

proving that

$$u(x,t) \leq \min_{(x(\cdot),\alpha)\in\mathcal{A}(x,t)} \left\{ u(x(r), r) + \int_t^r \frac{1}{2}|\alpha(s)|^2 + f(x(s), s)\, ds \right\}.$$

The opposite inequality follows from (6.1). $\square$

6.2. $\mathcal{G}$-differentials in $\mathbb{R}^2$

In this section, we introduce the notion of $\mathcal{G}$-differentiability in the 2-dimensional case and we collect several properties of semiconcave functions.

Definition 6.2. A function $u : \mathbb{R}^2 \to \mathbb{R}$ is $\mathcal{G}$-differentiable in $x \in \mathbb{R}^2$ if there exists $p_{\mathcal{G}} \in \mathbb{R}^2$ such that

$$\lim_{v\to 0} \frac{u(x_1 + v_1, x_2 + h(x_1)v_2) - u(x_1, x_2) - (p_{\mathcal{G}}, v)}{|v|} = 0;$$

in this case we denote $p_{\mathcal{G}} = D_{\mathcal{G}} u(x)$. We define the $\mathcal{G}$-subdifferential

$$D^-_{\mathcal{G}} u(x) := \{p \in \mathbb{R}^2 | \liminf_{v\to 0} \frac{u(x_1 + v_1, x_2 + h(x_1)v_2) - u(x_1, x_2) - (p, v)}{|v|} \geq 0\},$$

the lower $\mathcal{G}$-Dini derivative in the direction θ (i.e., $|\theta| = 1$)

$$\partial^-_{\mathcal{G}} u(x, \theta) := \liminf_{l\to 0^+, \theta'\to\theta} \frac{u(x_1 + l\theta'_1, x_2 + h(x_1)l\theta'_2) - u(x_1, x_2)}{l}$$

and the generalized $\mathcal{G}$-lower derivative in the direction θ

$$u^0_{\mathcal{G},-}(x, \theta) := \liminf_{l\to 0^+, y\to x} \frac{u(y_1 + l\theta_1, y_2 + h(y_1)l\theta_2) - u(y_1, y_2)}{l}.$$

The $\mathcal{G}$-superdifferential $D^+_{\mathcal{G}} u(x)$, the upper $\mathcal{G}$-Dini derivative $\partial^+_{\mathcal{G}} u(x, \theta)$ and the generalized $\mathcal{G}$-upper derivative $u^0_{\mathcal{G},+}(x, \theta)$ are defined in an analogous way. We introduce the reachable $\mathcal{G}$-gradients

$$D^*_{\mathcal{G}} u(x) := \{p : \exists x_n \to x,\ u \text{ is } \mathcal{G}\text{-differentiable at } x_n \text{ and } D_{\mathcal{G}} u(x_n) \to p\}.$$

We define the (1-sided) $\mathcal{G}$-directional derivative of u at x in the direction θ as

$$\partial_{\mathcal{G}} u(x,\theta) := \lim_{l\to 0^+} \frac{u(x_1+l\theta_1, x_2+h(x_1)l\theta_2) - u(x_1,x_2)}{l}.$$

Lemma 6.1.

1. *If u is $\mathcal{G}$-differentiable at x, then $D_{\mathcal{G}}u(x)$ is unique and $D^+_{\mathcal{G}}u(x)$ and $D^-_{\mathcal{G}}u(x)$ are both nonempty.*
2. *For $h(x_1) \neq 0$, there holds: $(p_1,p_2) \in D^+u(x)$ if and only if $(p_1, h(x_1)p_2) \in D^+_{\mathcal{G}}u(x)$.*
3. *For $h(x_1) = 0$ and $|\theta| = 1$, there holds:*

$$D^+_{\mathcal{G}}u(x) = \{(p_1, 0) : \limsup_{v_1\to 0} \frac{u(x_1+v_1,x_2) - u(x_1,x_2) - p_1 v_1}{|v_1|} \le 0\}$$

$$\partial_{\mathcal{G}} u(x,\theta) = \begin{cases} 0 & \textit{for } \theta_1 = 0 \\ |\theta_1|\partial u(x, (sgn(\theta_1), 0)) & \textit{for } \theta_1 \neq 0 \end{cases}$$

where $\partial u(x,\theta)$ is the standard directional derivative of u at x in the direction θ.

4. *For Lipschitz continuous function u, there holds:*

$$\partial^-_{\mathcal{G}} u(x,\theta) := \liminf_{l\to 0^+} \frac{u(x_1+l\theta_1, x_2+h(x_1)l\theta_2) - u(x_1,x_2)}{l}, \tag{6.4}$$

$$\textit{If } h(x_1) = 0 \textit{ then} \qquad (p_1,p_2) \in D^*_{\mathcal{G}}u(x) \Rightarrow p_2 = 0. \tag{6.5}$$

Proof. Points 1, 2 and 3 are obvious. The equality in (6.4) follows by the arguments of [11, Remark 3.1.4]. Let us prove (6.5). For any $(p_1,p_2) \in D^*_{\mathcal{G}}u(x)$, there exists $\{x_k\}_k$ with $x_k := (x_{k,1}, x_{k,2}) \to x$ and $D_{\mathcal{G}}u(x_k) \to (p_1,p_2)$. Possibly passing to a subsequence, we may assume that either $h(x_{k,1}) \neq 0$ for any k or $h(x_{k,1}) = 0$ for any k. In the first case, by Point 2, we have $D_{\mathcal{G}}u(x_k) = (D_1u(x_k), h(x_{k,1})D_2u(x_k))$ where D_1 and D_2 are the partial derivatives with respect to x_1 and x_2. As $k \to +\infty$, by the Lipschitz continuity of u, we get $p_2 = \lim_k h(x_{k,1})D_2u(x_k) = 0$. In the latter case, $D_{\mathcal{G}}u(x) = (D_1u(x_{k,1}), 0) \to (p_1, 0)$, the conclusion follows. □

Proposition 6.2. *We have*

$$D^+_{\mathcal{G}}u(x) = \{p : \partial^+_{\mathcal{G}}u(x,\theta) \le (p,\theta)\, \forall\theta \in \mathbb{R}^2\}, \quad D^-_{\mathcal{G}}u(x) = \{p : \partial^-_{\mathcal{G}}u(x,\theta) \ge (p,\theta)\, \forall\theta \in \mathbb{R}^2\}.$$

Moreover, $D^+_{\mathcal{G}}u(x)$ and $D^-_{\mathcal{G}}u(x)$ are both nonempty if and only if u is $\mathcal{G}$-differentiable at x and in this case they reduce to the singleton $D_{\mathcal{G}}u(x) = D^-_{\mathcal{G}}u(x) = D^+_{\mathcal{G}}u(x)$.

The proof of this proposition follows the same arguments of [11, Proposition 3.1.5]; actually the main difference is that one has to consider $x_k = (x_1 + v_{k,1}, x_2 + h(x_1)v_{k,2})$ with $v_k = (v_{k,1}, v_{k,2}) \to 0$. Hence we shall omit it.

We now generalize the Definition 2.3 of semiconcavity:

Definition 6.3. Let $u : \mathbb{R}^d \to \mathbb{R}$. We say that u is *semiconcave* if there exists a nondecreasing upper semicontinuous function $\omega : (0, +\infty) \to (0, +\infty)$ such that $\lim_{\rho\to 0^+} \omega(\rho) = 0$ and for all $\lambda \in [0, 1]$, for all $x, y \in \mathbb{R}^d$

$$\lambda u(y) + (1-\lambda)u(x) - 2u(\lambda y + (1-\lambda)x) \le \lambda(1-\lambda)|y-x|\omega(|y-x|).$$

The function ω is called modulus of semiconcavity of u.

Proposition 6.3. *Let u be a semiconcave function with modulus of semiconcavity ω. Then there hold*

1. *$p \in D^+_{\mathcal{G}}u(x)$ if and only if for any $v = (v_1, v_2) \in \mathbb{R}^2$*

$$u(x_1 + v_1, x_2 + h(x_1)v_2) - u(x_1, x_2) - (p, v) \le |(v_1, h(x_1)v_2)|\omega(|(v_1, h(x_1)v_2)|); \quad (6.6)$$

2. *If $\lim_k x_k = x$ and $p_k \in D^+_{\mathcal{G}}u(x_k)$ with $\lim_k p_k = p$, then $p \in D^+_{\mathcal{G}}u(x)$; hence, $D^*_{\mathcal{G}}u(x) \subset D^+_{\mathcal{G}}u(x)$;*
3. *$D^+_{\mathcal{G}}u(x, t) \ne \emptyset$;*
4. *If $D^+_{\mathcal{G}}u(x) = \{p\}$ (i.e., it is a singleton), then u is $\mathcal{G}$-differentiable at x.*

Proof. 1. Consider $p \in D^+_{\mathcal{G}}u(x)$. When $h(x_1) = 0$ and $v_1 = 0$, inequality (6.6) is a trivial consequence of Point 3 of Lemma 6.1. Otherwise, the rest of the proof is an adaptation of the argument in [11, Proposition 3.3.1] using [11, equation (2.1)] with $y = (x_1 + v_1, x_2 + h(x_1)v_2)$.

2. It follows directly from (6.6).

3. Being semiconcave, the function u is locally Lipschitz continuous. By Rademacher's theorem, there exists a sequence of points $\{x_k\}_k$ with $\lim_k x_k = x$ where u is differentiable and, in particular, $\mathcal{G}$-differentiable with $|D_{\mathcal{G}}u(x_k)| \le L$ (for some L). Possibly passing to a subsequence, $D_{\mathcal{G}}u(x_k) \to p$; hence, by point (2), $p \in D^+_{\mathcal{G}}u(x)$.

4. By Proposition 6.2, it suffices to prove: $p \in D^-_{\mathcal{G}}u(x)$. To this end, consider any sequence $\{v_k\}_k$, with $v_k \to 0$ as $k \to +\infty$ and introduce $\{x_k\}_k$ as

$$x_k = (x_{k,1}, x_{k,2}) := (x_1 + v_{k,1}, x_2 + h(x_{k,1})v_{k,2}).$$

We observe that: (i) $x_k \to x$ as $k \to +\infty$, (ii) by point (3), $\exists p_k \in D^+_{\mathcal{G}}u(x_k)$ with $|p_k| \le L$, (iii) by point (2) and possibly passing to a subsequence, $p_k \to p$ as $k \to +\infty$. Relation (6.6) centered in x_k defined above, with $v = -v_k$, gives

$$\begin{aligned} &- u(x_{k,1} - v_{k,1}, x_{k,2} - h(x_{k,1})v_{k,2}) + u(x_{k,1}, x_{k,2}) - (p_k, v_k) \\ &\qquad\qquad \ge -|(v_{k,1}, h(x_{k,1})v_{k,2})|\omega(|(v_{k,1}, h(x_{k,1})v_{k,2})|). \end{aligned}$$

By our choice of x_k, this inequality entails

$$\begin{aligned}
&\frac{-u(x_1,x_2)+u(x_1+v_{k,1},x_2+h(x_1)v_{k,2})-(p,v_k)}{|v_k|}\\
&\geq \frac{u(x_1+v_{k,1},x_2+h(x_1)v_{k,2})-u(x_1+v_{k,1},x_2+h(x_{k,1})v_{k,2})+(p_k-p,v_k)}{|v_k|}\\
&\quad -\frac{|(v_{k,1},h(x_{k,1})v_{k,2})|\omega(|(v_{k,1},h(x_{k,1})v_{k,2})|)}{|v_k|}\\
&\geq \frac{LL'|v_{k,2}||v_{k,1}|}{|v_k|}+(p_k-p,v_k/|v_k|)-\frac{|(v_{k,1},h(x_{k,1})v_{k,2})|\omega(|(v_{k,1},h(x_{k,1})v_{k,2})|)}{|v_k|}
\end{aligned}$$

where L and L' are respectively local Lipschitz constants of u and of h. Letting $k\to+\infty$, we obtain

$$\liminf_{k\to+\infty}\frac{u(x_1+v_{k,1},x_2+h(x_1)v_{k,2})-u(x_1,x_2)-(p,v_k)}{|v_k|}\geq 0;$$

by the arbitrariness of v_k, we conclude: $p\in D^-_{\mathcal{G}}u(x)$. □

In the next statement we establish that semiconcave functions always have directional derivatives.

Proposition 6.4. *Let u be a semiconcave function with modulus of semiconcavity ω. Then, for any direction θ, the directional derivative $\partial_{\mathcal{G}}u(x,\theta)$ exists and the following equalities hold:*

$$\partial_{\mathcal{G}}u(x,\theta)=\partial^-_{\mathcal{G}}u(x,\theta)=\partial^+_{\mathcal{G}}u(x,\theta)=u^0_{\mathcal{G},-}(x,\theta).$$

Proof. The proof is similar to the proof of [11, Theorem 3.2.1] so we just sketch it. Fix a direction θ and consider $0<l_1<l_2$. Relation [11, eq. (2.1)] with $\lambda=1-l_1/l_2$, $y=(x_1+l_2\theta_1,x_2+h(x_1)l_2\theta_2)$ entails

$$\begin{aligned}
\frac{u(x_1+l_1\theta_1,x_2+h(x_1)l_1\theta_2)-u(x)}{l_1}&\geq\frac{u(x_1+l_2\theta_1,x_2+h(x_1)l_2\theta_2)-u(x)}{l_2}\\
&\quad-\left(1-\frac{l_1}{l_2}\right)|(\theta_1,h(x_1)\theta_2)|\omega(l_2|(\theta_1,h(x_1)\theta_2)|). \qquad (6.7)
\end{aligned}$$

Passing to the $\liminf_{l_1\to0^+}$ and after to the $\limsup_{l_2\to0^+}$, we get $\partial^-_{\mathcal{G}}u(x,\theta)\geq\partial^+_{\mathcal{G}}u(x,\theta)$; hence, $\partial_{\mathcal{G}}u(x,\theta)$ exists and it coincides both with the upper and the lower $\mathcal{G}$-Dini derivatives. Moreover, by the definitions of $\partial^+_{\mathcal{G}}u(x,\theta)$ and of $u^0_{\mathcal{G},-}(x,\theta)$, Point 4 of Lemma 6.1 easily entails: $\partial^+_{\mathcal{G}}u(x,\theta)\geq u^0_{\mathcal{G},-}(x,\theta)$. Therefore, it remains to prove

$$\partial^+_{\mathcal{G}}u(x,\theta)\leq u^0_{\mathcal{G},-}(x,\theta). \qquad (6.8)$$

Let ϵ and $\overline{\ell}$ be two fixed positive constants with $\overline{\ell}\geq l$. Since u is continuous, there exists α sufficiently small such that

$$\frac{u(x_1+\overline{\ell}\theta_1,x_2+\overline{\ell}\theta_2h(x_1))-u(x)}{\overline{\ell}}\leq\frac{u(y_1+\overline{\ell}\theta_1,y_2+\overline{\ell}\theta_2h(y_1))-u(y)}{\overline{\ell}}+\epsilon \qquad \forall y\in B_\alpha(x).$$

By inequality (6.7) (with x, l_1 and l_2 replaced respectively by y, l), we get

$$\frac{u(y_1+\bar{\ell}\theta_1, y_2+\bar{\ell}\theta_2 h(y_1))-u(y)}{\bar{\ell}} \le \frac{u(y_1+l\theta_1, y_2+l\theta_2 h(y_1))-u(y)}{l} + \frac{\bar{\ell}-l}{\bar{\ell}}|(\theta_1, h(y_1)\theta_2)|\omega(\bar{\ell}|(\theta_1, h(y_1)\theta_2)|) \qquad \forall l\in(0,\bar{\ell}).$$

By the last two inequalities we deduce

$$\frac{u(x_1+\bar{\ell}\theta_1, x_2+\bar{\ell}\theta_2 h(x_1))-u(x)}{\bar{\ell}} \le \min_{y\in B_\alpha(x), l\in(0,\bar{\ell})} \frac{u(y_1+l\theta_1, y_2+l\theta_2 h(y_1))-u(y)}{l} + |(\theta_1, h(y_1)\theta_2)|\omega(\bar{\ell}|(\theta_1, h(y_1)\theta_2)|)+\epsilon.$$

Taking into account the definition of $u^0_{\mathcal{G},-}(x,\theta)$, we get

$$\frac{u(x_1+\bar{\ell}\theta_1, x_2+\bar{\ell}\theta_2 h(x_1))-u(x)}{\bar{\ell}} \le u^0_{\mathcal{G},-}(x,\theta)+|(\theta_1, h(x_1)\theta_2)|\omega(\bar{\ell}|(\theta_1, h(x_1)\theta_2)|)+\epsilon.$$

In conclusion, passing to the limit for $\epsilon\to 0^+$ and then $\limsup_{\bar{\ell}\to 0}$, we obtain inequality (6.8). □

Theorem 6.1. *Let u be a semiconcave function. Then, there holds*

$$D^+_{\mathcal{G}}u(x) = coD^*_{\mathcal{G}}u(x); \tag{6.9}$$

moreover, for any direction θ, the $\mathcal{G}$-directional derivative of u in the direction θ satisfies

$$\partial_{\mathcal{G}}u(x,\theta) = \min_{p\in D^+_{\mathcal{G}}u(x)}(p,\theta) = \min_{p\in D^*_{\mathcal{G}}u(x)}(p,\theta). \tag{6.10}$$

Proof. We shall use some of the arguments of [11, Theorem 3.3.6]. Let us prove relations (6.10). For any direction θ, using Proposition 6.2 and Proposition 6.3-(2), we obtain

$$\partial_{\mathcal{G}}u(x,\theta) \le \min_{p\in D^+_{\mathcal{G}}u(x)}(p,\theta) \le \min_{p\in D^*_{\mathcal{G}}u(x)}(p,\theta).$$

Hence, it remains to prove

$$\min_{p\in D^*_{\mathcal{G}}u(x)}(p,\theta) \le \partial_{\mathcal{G}}u(x,\theta) \qquad \text{for any direction } \theta. \tag{6.11}$$

In order to prove this inequality, we study separately the cases when x_1 belongs or not to $\{h(x_1)=0\}$. Assume $h(x_1)\neq 0$ and fix a direction θ. Since u is differentiable a.e., there exists a sequence $\{v_k\}_k$, with $v_k\in\mathbb{R}^2$, such that: (i) $v_k\to 0$ as $k\to+\infty$, (ii) $v_k/|v_k|\to\theta$ as $k\to+\infty$, (iii) u is differentiable at $x_k := (x_1+v_{k,1}, x_2+v_{k,2}h(x_1))$, (iv) (taking advantage of the Lipschitz continuity of u and possibly passing to a subsequence) $D_{\mathcal{G}}u(x_k)$ converge to some $p \in D^*_{\mathcal{G}}u(x)$ as $k \to +\infty$. Applying inequality [11, eq. (3.18)] (with x and y replaced respectively by x_k and x), we get

$$u(x) - u(x_k) + (Du(x_k), (v_{k,1}, h(x_1)v_{k,2})) \leq |(v_{k,1}, h(x_1)v_{k,2})|\omega(|(v_{k,1}, h(x_1)v_{k,2})|). \quad (6.12)$$

On the other hand, we observe that point (iii) here above and Point 2 of Lemma 6.1 ensure that u is $\mathcal{G}$-differentiable at x_k with $D_{\mathcal{G}}u(x_k) = (D_1u(x_k), h(x_{k,1})D_2u(x_k))$. Hence, we have

$$\begin{aligned}(Du(x_k), (v_{k,1}, h(x_1)v_{k,2})) &= (D_{\mathcal{G}}u(x_k), v_k) + D_2u(x_k)v_{k,2}[h(x_1) - h(x_{k,1})] \\ &\geq (D_{\mathcal{G}}u(x_k), v_k) - C|v_{k,2}||v_{k,1}|,\end{aligned} \quad (6.13)$$

where the last inequality holds for a suitable $C > 0$, and is due to the Lipschitz continuity of u and of h. By (6.12) and (6.13), we get

$$(D_{\mathcal{G}}u(x_k), v_k/|v_k|) \leq \frac{u(x_k) - u(x)}{|v_k|} + \frac{C|v_{k,2}||v_{k,1}|}{|v_k|} + \frac{|(v_{k,1}, h(x_1)v_{k,2})|}{|v_k|}\omega(|(v_{k,1}, h(x_1)v_{k,2})|).$$

Letting $k \to +\infty$, we infer: $(p, \theta) \leq \partial_{\mathcal{G}}u(x, \theta)$ for some $p \in D^*_{\mathcal{G}}u(x)$ which, in turns, entails (6.11).

Consider now x such that $h(x_1) = 0$. By Point 4 of Lemma 6.1 we have: $\min_{p \in D^*_{\mathcal{G}}u(x)}(p, \theta) = \min_{p \in D^*_{\mathcal{G}}u(x)} p_1\theta_1$; taking into account also Point 3 of Lemma 6.1, relation (6.11) is equivalent to

$$\min_{p \in D^*_{\mathcal{G}}u(x)} p_1 \mathrm{sgn}(\theta_1) \leq \partial u(x, (\mathrm{sgn}(\theta_1), 0)) \qquad \forall \theta_1 \in [-1, 1] \setminus \{0\}.$$

In order to prove this relation, we follow an argument similar to the previous case. We consider a sequence $\{v_k\}_k$ such that: (i) $v_k \to 0$ as $k \to +\infty$, (ii) $v_k/|v_k| \to (\mathrm{sgn}(\theta_1), 0)$ as $k \to +\infty$ (in particular $v_{k,2}/|v_k| \to 0$), (iii) u is differentiable at $x_k := (x_1 + v_{k,1}, x_2 + v_{k,2})$ (note that this definition is different from the corresponding one in the previous case), (iv) $D_{\mathcal{G}}u(x_k)$ converge to some $p \in D^*_{\mathcal{G}}u(x)$ as $k \to +\infty$. Applying inequality [11, eq. (3.18)] (with x and y replaced respectively by x_k and x), we get

$$(Du(x_k), v_k) \leq u(x_k) - u(x) + |v_k|\omega(|v_k|).$$

Again we get that u is $\mathcal{G}$-differentiable at x_k with $D_u(x_k) = (D_1u(x_k), h(x_{k,1})D_2u(x_k))$. Hence, we deduce

$$(Du(x_k), v_k) = (D_{\mathcal{G}}u(x_k), v_k) + D_2u(x_k)[1 - h(x_{k,1})]v_{k,2} \geq (D_{\mathcal{G}}u(x_k), v_k) - C|v_{k,2}|$$

where the last inequality is due to the Lipschitz continuity of u and to the boundedness of h. By the last two inequalities, we get

$$(D_{\mathcal{G}}u(x_k), v_k/|v_k|) \leq \frac{u(x_k) - u(x)}{|v_k|} + \frac{C|v_{k,2}|}{|v_k|} + \omega(|v_k|).$$

Letting $k \to +\infty$, we infer: $p_1\mathrm{sgn}(\theta_1) \leq \partial u(x, (\mathrm{sgn}(\theta_1), 0))$. Hence, relations (6.10) are completely proved. Arguing as in [11, Theorem 3.3.6], we infer relation (6.9). $\square$

6.3. $\mathcal{G}$-differentials in $\mathbb{R}^d$

In this subsection we extend the definition of $\mathcal{G}$-differentiability in the d-dimensional case which is used along Section 5.

Definition 6.4. A function $u : \mathbb{R}^d \to \mathbb{R}$ is $\mathcal{G}$-differentiable in $x \in \mathbb{R}^d$ if there exists $\rho_{\mathcal{G}} \in \mathbb{R}^d$ such that

$$\lim_{v \to 0} \frac{u(\tilde{x}) - u(x) - (\rho_{\mathcal{G}}, v)}{|v|} = 0;$$

where, for $v \in \mathbb{R}^d$ we iteratively define $\tilde{x}_1 = x_1 + h_{11} v_1$, and $\tilde{x}_i = x_i + \sum_{j=1}^d h_{ij}(\tilde{x}_1, \cdots, \tilde{x}_{i-1}) v_j$, where h_{ij} are defined in (5.3).

In a way similar to Definition 6.2, we define the $\mathcal{G}$-subdifferential and $\mathcal{G}$-superdifferential, the lower and upper $\mathcal{G}$-Dini derivative in the direction $\theta \in \mathbb{R}^d$, the generalized $\mathcal{G}$-upper and $\mathcal{G}$-lower derivative in the direction θ, the reachable $\mathcal{G}$-gradients, the (1-sided) $\mathcal{G}$-directional derivative of u at x in the direction θ.

Remark 6.1. If u is differentiable then $\rho_{\mathcal{G}} =: D_{\mathcal{G}} u = Du\, B$, where B is defined in (5.3).

By the same arguments as in the 2-dim case in Section 6.2, we get the following results.

Proposition 6.5. *Let u be a semiconcave function with modulus of semiconcavity ω. Then there hold*

1. $p \in D^+_{\mathcal{G}} u(x)$ *if and only if for any* $v \in \mathbb{R}^d$

$$u(\tilde{x}) - u(x) - (\rho_{\mathcal{G}}, v) \le |\tilde{x} - x| \omega(|\tilde{x} - x|), \tag{6.14}$$

   where $\tilde{x}$ is defined as in Definition 6.4.
2. *If* $\lim_k x_k = x$ *and* $p_k \in D^+_{\mathcal{G}} u(x_k)$ *with* $\lim_k p_k = p$, *then* $p \in D^+_{\mathcal{G}} u(x)$; *hence,* $D^*_{\mathcal{G}} u(x) \subset D^+_{\mathcal{G}} u(x)$;
3. $D^+_{\mathcal{G}} u(x, t) \neq \emptyset$;
4. *If* $D^+_{\mathcal{G}} u(x) = \{p\}$ *(i.e., it is a singleton), then u is $\mathcal{G}$-differentiable at x.*

Theorem 6.2. *Let u be a semiconcave function. Then, there holds*

$$D^+_{\mathcal{G}} u(x) = co D^*_{\mathcal{G}} u(x); \tag{6.15}$$

moreover, for any direction θ, the $\mathcal{G}$-directional derivative of u in the direction θ satisfies

$$\partial_{\mathcal{G}} u(x, \theta) = \min_{p \in D^+_{\mathcal{G}} u(x)} (p, \theta) = \min_{p \in D^*_{\mathcal{G}} u(x)} (p, \theta). \tag{6.16}$$

Acknowledgments

The authors wish to warmly thank P. Cardaliaguet for helpful discussions and hints. They also wish to warmly thank the anonymous referee for useful suggestions and comments to improve the paper. The first and the second authors are members of GNAMPA-INdAM and were partially supported also by the research project Prot. CPDA157835 of the University of Padova "Mean-Field Games and Nonlinear PDEs" and by the Fondazione CaRiPaRo Project "Nonlinear Partial Differential Equations: Asymptotic Problems and Mean-Field Games". The third author was partially supported by the Padova University grant SID 2018 "Controllability, stabilizability and infimum gaps for control systems", prot. BIRD 187147. The fourth author has been partially funded by the ANR project ANR-16-CE40-0015-01.

References

[1] Y. Achdou, I. Capuzzo Dolcetta, Mean field games: numerical methods, SIAM J. Numer. Anal. 48 (2010) 1136–1162.
[2] Y. Achdou, P. Mannucci, C. Marchi, N. Tchou, Deterministic mean field games with control on the acceleration, available at https://arxiv.org/abs/1908.03330.
[3] L. Ambrosio, N. Gigli, G. Savaré, Gradient Flows in Metric Spaces and in the Space of Probability Measures, Lectures in Mathematics ETH Zürich, Birkhäuser Verlag, Basel, 2005.
[4] M. Bardi, I. Capuzzo Dolcetta, Optimal Control and Viscosity Solutions of Hamilton-Jacobi Bellman Equations, Systems and Control: Foundations and Applications, Birkhäuser, Boston, 1997.
[5] M. Bardi, F. Dragoni, Subdifferential and properties of convex functions with respect to vector fields, J. Convex Anal. 21 (3) (2014) 785–810.
[6] J.-D. Benamou, G. Carlier, F. Santambrogio, Variational mean field games, in: Active Particles. Vol. 1. Advances in Theory, Models, and Applications, Birkhäuser/Springer, Cham, 2017, pp. 141–171.
[7] A. Bensoussan, J. Frehse, P. Yam, Mean Field Games and Mean Field Type Control Theory, Springer Briefs in Mathematics, Springer, New York, 2013.
[8] B. Bonnard, M. Chyba, Singular Trajectories and Their Role in Control Theory, Mathematics & Applications, vol. 40, Springer-Verlag, Berlin, 2003.
[9] P. Cannarsa, R. Capuani, Existence and uniqueness for Mean Field Games with state constraints, in: P. Cardaliaguet, A. Porretta, F. Salvarani (Eds.), PDE Models for Multi-Agent Phenomena, in: Springer INdAM Ser., vol. 28, Springer, Cham, 2018, pp. 49–71.
[10] P. Cannarsa, L. Rifford, Semiconcavity results for optimal control problems admitting no singular minimizing controls, Ann. Inst. Henri Poincaré, Anal. Non Linéaire 25 (4) (2008) 773–802.
[11] P. Cannarsa, C. Sinestrari, Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control, Progress in Nonlinear Differential Equations and Their Applications, vol. 48, Birkhäuser, Boston, 2004.
[12] P. Cardaliaguet, Notes on Mean Field Games, 2012, from P.L. Lions lectures at College de France, available at https://www.ceremade.dauphine.fr/cardalia/MFG20130420.pdf.
[13] P. Cardaliaguet, Long time average of first order mean field games and weak KAM theory, Dyn. Games Appl. 3 (2013) 473–488.
[14] P. Cardaliaguet, P.J. Graber, A. Porretta, D. Tonon, Second order mean field games with degenerate diffusion and local coupling, NoDEA Nonlinear Differ. Equ. Appl. 22 (2015) 1287–1317.
[15] P. Cardaliaguet, S. Hadikhanloo, Learning in mean field games: the fictitious play, ESAIM Control Optim. Calc. Var. 23 (2) (2017) 569–591.
[16] F. Clarke, Functional Analysis, Calculus of Variations and Optimal Control, Graduate Text in Mathematics, vol. 264, Springer-Verlag, London, 2013.
[17] F. Clarke, Optimization and Nonsmooth Analysis, 2nd edition, Classics in Applied Mathematics, vol. 5, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1990.
[18] F. Dragoni, E. Feleqi, Ergodic Mean Field Games with Hörmander diffusions, Calc. Var. Partial Differ. Equ. 57 (5) (2018) 116.
[19] D. Gomes, E.A. Pimentel, V. Voskanyan, Regularity Theory for Mean-Field Game Systems, Springer Briefs in Mathematics, Springer, Berlin, 2016.
[20] D. Gomes, J. Saude, Mean field games - a brief survey, Dyn. Games Appl. 4 (2014) 110–154.

[21] M. Huang, R.P. Malhamé, P.E. Caines, Large population stochastic dynamic games: closed-loop McKean-Vlasov systems and the Nash certainty equivalence principle, Commun. Inf. Syst. 6 (2006) 221–251.
[22] I. Karatzas, S.E. Shreve, Brownian Motion and Stochastic Calculus, second edition, Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991.
[23] V.I. Bogachev, N.V. Krylov, M. Röckner, S.V. Shaposhnikov, Fokker-Planck-Kolmogorov Equations, Mathematical Surveys and Monographs, vol. 207, American Mathematical Society, Providence, RI, 2015.
[24] J.-M. Lasry, P.-L. Lions, Jeux à champ moyen. I. Le cas stationnaire, C. R. Math. Acad. Sci. Paris 343 (2006) 619–625.
[25] J.-M. Lasry, P.-L. Lions, Jeux à champ moyen. II. Horizon fini et contrôle optimal, C. R. Math. Acad. Sci. Paris 343 (2006) 679–684.
[26] J.-M. Lasry, P.-L. Lions, Mean field games, Jpn. J. Math. (N.S.) 2 (2007) 229–260.
[27] G.M. Lieberman, Second Order Parabolic Differential Equations, World Scientific Publishing Co., Inc., River Edge, NJ, 1996.
[28] P.L. Lions, M. Musiela, Ergodicity of diffusion processes, unpublished.