



Non coercive unbounded first order Mean Field Games: The Heisenberg example

Paola Mannucci ^a, Claudio Marchi ^b, Nicoletta Tchou ^c

^a *Dipartimento di Matematica “Tullio Levi-Civita”, Università di Padova, Italy*

^b *Dipartimento di Ingegneria dell’Informazione & Dipartimento di Matematica “Tullio Levi-Civita”, Università di Padova, Italy*

^c *Univ Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France*

Received 22 February 2021; revised 22 October 2021; accepted 20 November 2021

Available online 7 December 2021

Abstract

In this paper we study evolutive first order Mean Field Games in the Heisenberg group; each agent can move in the whole space but it has to follow “horizontal” trajectories which are given in terms of the vector fields generating the group and the kinetic part of the cost depends only on the horizontal velocity. The Hamiltonian is not coercive in the gradient term and the coefficients of the first order term in the continuity equation may have a quadratic growth at infinity. The main results of this paper are two: the former is to establish the existence of a weak solution to the Mean Field Game systems while the latter is to represent this solution following the Lagrangian formulation of the Mean Field Games. We also provide some generalizations to Heisenberg-type structures.

© 2021 Elsevier Inc. All rights reserved.

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

Keywords: Mean Field Games; First order Hamilton-Jacobi equations; Continuity equation; Fokker-Planck equation; Noncoercive Hamiltonian; Heisenberg group; Heisenberg-type groups; Degenerate optimal control problem

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

<https://doi.org/10.1016/j.jde.2021.11.029>

0022-0396/© 2021 Elsevier Inc. All rights reserved.

1. Introduction

In this paper we study evolutive first order Mean Field Game (briefly, MFG) systems in the Heisenberg group \mathbb{H}^1 . Let us recall that the MFG theory started with the works by Lasry and Lions [29–31] and by Huang, Malhamé and Caines [26] (see the works [16,8,25] for the many developments in recent years) and studies Nash equilibria when the number of agents tends to infinity and each agent’s aim is to control its dynamics so to minimize a given cost which depends on the distribution of the whole population. On the other hand, the Heisenberg group can be seen as the first non-Euclidean space which is still endowed with nice properties as a (noncommutative) group operation, a family of dilations and a manifold structure (see the monographs [11,34] for an overview). The Heisenberg setting gives rise to non holonomic constraints in the optimal problem addressed by the agent, see [20, p. 135] and [6, p. 52]. Moreover let us recall that many control problems, as the vehicle model, can be written in a form similar to the Heisenberg one: for real models and numerical analysis see [1,35,36,22] and references therein. From the viewpoint of a single agent in the MFG, the Heisenberg’s framework entails that its state cannot change isotropically in all the directions but it can move only along *admissible* trajectories.

We shall consider systems of the form

$$\begin{cases} (i) & -\partial_t u + \frac{|D_{\mathcal{H}}u|^2}{2} = F[m(t)](x) & \text{in } \mathbb{H}^1 \times (0, T) \\ (ii) & \partial_t m - \operatorname{div}_{\mathcal{H}}(m D_{\mathcal{H}}u) = 0 & \text{in } \mathbb{H}^1 \times (0, T) \\ (iii) & m(x, 0) = m_0(x), u(x, T) = G[m(T)](x) & \text{on } \mathbb{H}^1, \end{cases} \tag{1.1}$$

where $D_{\mathcal{H}}$ and $\operatorname{div}_{\mathcal{H}}$ are respectively the *horizontal gradient* and the *horizontal divergence* while F and G are strongly regularizing coupling operators. For readers which are not familiar with intrinsic calculus, in Euclidean coordinates, system (1.1) becomes

$$\begin{cases} (i) & -\partial_t u + H(x, Du) = F[m(t)](x) & \text{in } \mathbb{R}^3 \times (0, T) \\ (ii) & \partial_t m - \operatorname{div}(m \partial_p H(x, Du)) = 0 & \text{in } \mathbb{R}^3 \times (0, T) \\ (iii) & m(x, 0) = m_0(x), u(x, T) = G[m(T)](x) & \text{on } \mathbb{R}^3, \end{cases} \tag{1.2}$$

where, for $p = (p_1, p_2, p_3)$ and $x = (x_1, x_2, x_3)$, the Hamiltonian $H(x, p)$ is

$$H(x, p) := \frac{1}{2}((p_1 - x_2 p_3)^2 + (p_2 + x_1 p_3)^2) = \frac{|pB(x)|^2}{2}$$

$$\text{with } B(x) := \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -x_2 & x_1 \end{pmatrix} \in M^{3 \times 2} \tag{1.3}$$

while the drift $\partial_p H(x, p)$ is

$$\partial_p H(x, p) = pB(x)B(x)^T = (p_1 - x_2 p_3, p_2 + x_1 p_3, -p_1 x_2 + p_2 x_1 + p_3(x_1^2 + x_2^2)). \tag{1.4}$$

These MFG systems arise when the generic player with state x at time t can move in the whole space but it must follow *horizontal curves* with respect to the two vector fields X_1 and X_2 generating the Heisenberg group (see (2.1) below):

$$x'(s) = \alpha_1(s)X_1(x(s)) + \alpha_2(s)X_2(x(s)) \tag{1.5}$$

namely

$$x'_1(s) = \alpha_1(s), \quad x'_2(s) = \alpha_2(s), \quad x'_3(s) = -x_2(s)\alpha_1(s) + x_1(s)\alpha_2(s).$$

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

$$J^m_{x,t}(\alpha) := \int_t^T \left[\frac{1}{2} |\alpha(\tau)|^2 + F[m(\tau)](x(\tau)) \right] d\tau + G[m(T)](x(T)) \tag{1.6}$$

where $m(\cdot)$ is the evolution of the whole population’s distribution while $(x(\cdot), \alpha(\cdot))$ is a trajectory obeying to (1.5).

Let us observe three important issues of these MFG systems: (i) the Hamiltonian H is not coercive in p , (ii) the system is in the whole space, (iii) in equation (1.2)-(ii) the coefficient of the first order term may have quadratic growth in x .

Point (i) prevents the application of standard approaches for first order MFG (for instance, see [8,16,18]) because they require uniform coercivity of the Hamiltonian. Moreover, we recall that the papers [2,13,33] already tackled MFG systems with noncoercive Hamiltonians for first order MFG while papers [23,24] dealt with second order hypoelliptic MFG. However, the results in [2,13] do not apply to the present setting because these papers consider a different kind of admissible trajectories. Note that the present case is not even encompassed in the previous work [33] because $\det B(x)B(x)^T = 0$ for any $x \in \mathbb{R}^3$. The degeneracy of the matrix $B(x)B(x)^T$ implies that we cannot prove the uniqueness of optimal trajectories for a.e. starting points with respect to the initial distribution of players and hence to get a representation formula as in [16,33]. The issue of finding necessary or sufficient conditions ensuring the uniqueness of the optimal trajectories for a.e. starting points is challenging and open; we hope to study it in a future work.

On the other hand, points (ii) and (iii) give rise to some difficulties for applying the vanishing viscosity method, especially for the Cauchy problem for equation (1.2)-(ii) with the viscosity term. Actually in this problem the coefficients grow “too much at infinity” and one cannot invoke nor standard results for the well-posedness of the problem neither its interpretation in terms of a stochastic optimal control problem.

The aims of this paper are two; the former one is to prove the existence of a weak solution to system (1.1) while the latter, and main, one is to prove that this weak solution is also a *mild* solution in the sense introduced by Cannarsa and Capuani [12] for the case of state-constrained MFG where the agents control their velocity.

In order to obtain the existence of a weak solution, we establish several properties of the solution to the Hamilton-Jacobi equation (1.1)-(i) (as semiconcavity, Lipschitz continuity, regularity of the optimal trajectories for the associated optimal control problem). Afterwards, we adapt the techniques introduced by PL. Lions in his lectures at Collège de France [16,31] (see also [2,33] for similar approaches for some noncoercive Hamiltonians). To get the result we perform three approximations: a completion B^ε of B (see (2.8)), a vanishing viscosity procedure with the *Euclydean* Laplacian and a truncation argument of the coefficients of matrix B . The completion B^ε fulfills $\det B^\varepsilon(x)(B^\varepsilon(x))^T \neq 0$ for any $x \in \mathbb{R}^3$ which is a crucial property for getting uniqueness of optimal trajectory for m_0 -a.e. starting point. The vanishing viscosity procedure permits to exploit the regularity results of the Laplacian while the truncation argument permits to avoid parabolic Cauchy problems with coefficients growing “too much” at infinity.

Let us note that the matrix B^ε is associated to different constraints on horizontal curves (see (2.9) below) and that the geometry of the space changes drastically as $\varepsilon \rightarrow 0^+$; we refer the reader to the paper [15] for a discussion on this issue. Finally, we shall prove that this weak solution is also a *mild* solution in the sense introduced in [12]. Roughly speaking, as in the Lagrangian approach for MFG (see [7]), this property means that, for a.e. starting state, the agents follow optimal trajectories for the optimal control problem associated to the Hamilton-Jacobi equation. In order to prove that our solution is in fact a mild solution, we shall use the superposition principle [3, Theorem 8.2.1]. To do this we need to prove in the Heisenberg framework an optimal synthesis result and the aforementioned properties of the matrix B^ε .

It is worth noting that our techniques relies on some compactness of initial distribution of players and on sublinear growth of the coefficients of B but they do not need the Hörmander condition. Indeed, we present our results for purely quadratic Hamiltonian as in (1.3) on the (first) Heisenberg group \mathbb{H}^1 only for the sake of simplicity. As a matter of facts, our results can be extended to any Hamiltonian of the form $H(x, p) = |pB(x)|^\gamma$ with $\gamma \in [1, 2]$ and some structures of *Heisenberg type* (see [11, Chapter 18] for precise definition and main properties). As an example, we apply our result to the case of classical Grushin dynamics with unbounded coefficients. For the case of Grushin dynamics with bounded coefficients, we refer the reader to [33, Theorem 1.1] where a more precise interpretation of the system is obtained taking advantage of such a boundedness.

This paper is organized as follows. Section 2 is devoted to give the main definitions including the Heisenberg group, contains the assumptions and the statement of our main result (Theorem 2.1) whose proof is postponed in Section 5. In Section 3 we study several properties of the solution of the optimal control problem associated to the Hamilton-Jacobi equation (1.1)-(i). Section 4 is devoted to establish the well posedness of the continuity equation; in particular, Theorem 4.1 states some regularity estimates which are crucial in the proof of Theorem 2.1. In Section 6 we provide two generalizations of our result: in the former we consider structures of Heisenberg type while in the latter we tackle power-type Hamiltonians with exponent $\gamma \in [1, 2]$.

2. Preliminaries: definitions, assumptions and main results

In this section, we introduce the notations (including the functional spaces needed for the definition of solution to system (1.1) and the Heisenberg group), fix our assumptions and state the main results of this paper.

2.1. Notations and Heisenberg group

For any function $u : \mathbb{R}^n \times \mathbb{R} \ni (x, t) \rightarrow u(x, t) \in \mathbb{R}$, Du and D^2u stand for the Euclidean gradient and respectively Hessian matrix with respect to x . We denote $C^2(\mathbb{R}^n)$ the space of functions with continuous second order derivatives and we write $\|f\|_{C^2(\mathbb{R}^n)} := \sup_{x \in \mathbb{R}^n} [|f(x)| + |Df(x)| + |D^2f(x)|]$.

For any complete separable metric space X , $\mathcal{P}(X)$ denotes the set of Borel probability measures on X . For any complete separable metric spaces X_1 and X_2 , any measure $\eta \in \mathcal{P}(X_1)$ and any function $\phi : X_1 \rightarrow X_2$, we denote $\phi\#\eta \in \mathcal{P}(X_2)$ the *push-forward* of η through ϕ , i.e. $\phi\#\eta(B) := \eta(\phi^{-1}(B))$, for any B measurable set, $B \subset X_2$ (see [3, section 5.2] for its main properties). For a function $m \in C([0, T], \mathcal{P}(X))$, m_t stands for the probability $m(t, \cdot)$ on X .

We introduce the functional spaces $\mathcal{P}_1(\mathbb{R}^n)$ (respectively, $\mathcal{P}_2(\mathbb{R}^n)$) as the space of Borel probability measures on \mathbb{R}^n with finite first (resp., second) order moment with respect to the Euclidean

distance, endowed with the Monge-Kantorovich distance \mathbf{d}_1 (resp., \mathbf{d}_2) (for more details see [3] or [16]).

We refer to [11] for a complete overview on the Heisenberg group \mathbb{H}^1 . We define the two vector fields, called *generators* of \mathbb{H}^1 ,

$$X_1(x) := \begin{pmatrix} 1 \\ 0 \\ -x_2 \end{pmatrix} \quad \text{and} \quad X_2(x) := \begin{pmatrix} 0 \\ 1 \\ x_1 \end{pmatrix}, \quad \forall x = (x_1, x_2, x_3) \in \mathbb{R}^3. \tag{2.1}$$

By these vectors we define the linear differential operators, still called X_1 and X_2

$$X_1 = \partial_{x_1} - x_2 \partial_{x_3}, \quad X_2 = \partial_{x_2} + x_1 \partial_{x_3}. \tag{2.2}$$

Note that their *commutator* $[X_1, X_2] := X_1 X_2 - X_2 X_1$ verifies: $[X_1, X_2] = -2\partial_{x_3}$; hence, X_1 , X_2 and $[X_1, X_2]$ span all \mathbb{R}^3 .

The Heisenberg group \mathbb{H}^1 has a group structure endowed with the following noncommutative group operation, denoted by \oplus : for all $x = (x_1, x_2, x_3)$, $y = (y_1, y_2, y_3) \in \mathbb{R}^3$,

$$x \oplus y = (x_1, x_2, x_3) \oplus (y_1, y_2, y_3) := (x_1 + y_1, x_2 + y_2, x_3 + y_3 - x_2 y_1 + x_1 y_2).$$

The fields X_1 and X_2 are left-invariant vector fields, i.e. for all $u \in C^\infty(\mathbb{R}^3)$ and for all fixed $y \in \mathbb{R}^3$ we have $X_i(u(y \oplus x)) = (X_i u)(y \oplus x)$, $i = 1, 2$.

Note that the matrix $B(x)$ defined in (1.3) is the matrix associated to the vectors X_1 and X_2 . For any regular real-valued function u , we shall denote its horizontal gradient and its horizontal Laplacian by $D_{\mathcal{H}}u := (X_1 u, X_2 u)$ and respectively $\Delta_{\mathcal{H}} := X_1^2 u + X_2^2 u$ and we observe $D_{\mathcal{H}}u = Du B(x)$ and $\Delta_{\mathcal{H}}u = \text{tr}(D^2 u B B^T)$. For any regular $v = (v_1, v_2) : \mathbb{R}^3 \rightarrow \mathbb{R}^2$, we denote its horizontal divergence by $\text{div}_{\mathcal{H}} v := X_1 v_1 + X_2 v_2$ and we note that the left-invariance of X_i ($i = 1, 2$) entails the left-invariance of $\text{div}_{\mathcal{H}}$. We have: $\text{div}_{\mathcal{H}}(D_{\mathcal{H}}u) = \Delta_{\mathcal{H}}u$.

The norm and the distance associated by the group law are defined as

$$\|x\|_{\mathcal{H}} := ((x_1^2 + x_2^2)^2 + x_3^2)^{1/4}, \quad d_{\mathcal{H}}(x, y) := \|x \oplus y^{-1}\|_{\mathcal{H}}.$$

2.2. Assumptions, definitions of solution and main result

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 $\mathcal{P}_1(\mathbb{R}^3) \times \mathbb{R}^3$,
- (H2) the map $m \rightarrow F[m](\cdot)$ is Lipschitz continuous from $\mathcal{P}_1(\mathbb{R}^3)$ to $C^2(\mathbb{R}^3)$; moreover, there exists $C \in \mathbb{R}$ such that

$$\|F[m](\cdot)\|_{C^2(\mathbb{R}^3)}, \|G[m](\cdot)\|_{C^2(\mathbb{R}^3)} \leq C, \quad \forall m \in \mathcal{P}_1(\mathbb{R}^3);$$

- (H3) the distribution $m_0 : \mathbb{R}^3 \rightarrow \mathbb{R}$ is a nonnegative C^0 function with compact support and $\int_{\mathbb{R}^3} m_0 dx = 1$.

We now introduce our definitions of weak solution of the MFG system (1.1).

Definition 2.1. A couple (u, m) of functions defined on $\mathbb{R}^3 \times [0, T]$ is a weak solution of system (1.1) if

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

Remark 2.1. From [3, Lemma 8.1.2], we get that the distributional solution of (1.1)-(ii) stated in point 4) of the Definition 2.1 is automatically continuous in the sense of point 2) of the same definition.

We introduce now the notion of *mild* solution introduced by [12]. This notion is reminiscent of the Lagrangian approach to MFGs (see [7]) and it relies on replacing probability measures on the state space with probability measures on arcs on the state space.

We define the set of AC arcs in \mathbb{R}^3

$$\Gamma := AC((0, T), \mathbb{R}^3) \tag{2.3}$$

and the evaluation map $e_t : \Gamma \rightarrow \mathbb{R}^3$ as

$$e_t(\gamma) = \gamma(t). \tag{2.4}$$

For any $x \in \mathbb{R}^3$, we define the set of horizontal arcs starting at x with an associated control law

$$\mathcal{A}(x, t) := \{(\gamma, \alpha) : \gamma \in \Gamma, \gamma(t) = x, \alpha \in L^2([t, T], \mathbb{R}^2), (\gamma, \alpha) \text{ solves (1.5) in } (t, T)\}.$$

Given $m_0 \in \mathcal{P}_1(\mathbb{R}^3)$, we define

$$\mathcal{P}_{m_0}(\Gamma) = \{\eta \in \mathcal{P}(\Gamma) : m_0 = e_0\#\eta\}.$$

For any $\eta \in \mathcal{P}_{m_0}(\Gamma)$, $t \in [0, T]$ and $x \in \mathbb{R}^3$, we consider the cost

$$J_{x,t}^\eta(\alpha) := \int_t^T \left[\frac{1}{2} |\alpha(\tau)|^2 + F[e_\tau\#\eta](\gamma(\tau)) \right] d\tau + G[e_T\#\eta](\gamma(T)) \tag{2.5}$$

where $(\gamma, \alpha) \in \mathcal{A}(x, t)$. For any $\eta \in \mathcal{P}_{m_0}(\Gamma)$ and for any $x \in \mathbb{R}^3$ we define the set of optimal horizontal arcs starting at x

$$\Gamma^\eta[x] := \{\bar{\gamma} : (\bar{\gamma}, \bar{\alpha}) \in \mathcal{A}(x, 0) : J_{x,0}^\eta(\bar{\alpha}) = \min_{(\gamma,\alpha) \in \mathcal{A}(x,0)} J_{x,0}^\eta(\alpha)\}. \tag{2.6}$$

Definition 2.2. A measure $\eta \in \mathcal{P}_{m_0}(\Gamma)$ is a *MFG equilibrium* for m_0 if

$$\text{supp } \eta \subseteq \bigcup_{x \in \mathbb{R}^3} \Gamma^\eta[x].$$

This means that the support of η is contained in the set

$$\bigcup_{x \in \mathbb{R}^3} \{\gamma \in \Gamma : \gamma(0) = x, \gamma \text{ is a minimizer of } J_{x,0}^\eta\}.$$

Definition 2.3. A couple $(u, m) \in C^0([0, T] \times \mathbb{R}^3) \times C^0([0, T]; \mathcal{P}_1(\mathbb{R}^3))$ is called *mild solution* if there exists a MFG equilibrium η for m_0 such that:

- i) $m_t = e_t \# \eta$;
- ii) u is given by

$$u(x, t) = \inf_{(\gamma, \alpha) \in \mathcal{A}(x, t)} \int_t^T \left[\frac{1}{2} |\alpha(\tau)|^2 + F[e_\tau \# \eta](\gamma(\tau)) \right] d\tau + G[e_T \# \eta](\gamma(T)).$$

We now state the main result of this paper.

Theorem 2.1. *Under the above assumptions:*

- i) System (1.1) has a solution (u, m) ;
- ii) (u, m) is a mild solution.

Remark 2.2.

- As a matter of fact, in the proof of this theorem we get that any solution in the sense of Definition 2.1 is a mild solution.
- Uniqueness holds under classical hypothesis on the monotonicity of F and G as in [16].

2.3. The ε -approximating problem

In order to prove Theorem 2.1, it is expedient to introduce the following approximating problems for $\varepsilon \in (0, 1]$

$$\begin{cases} (i) & -\partial_t u + H^\varepsilon(x, Du) = F[m(t)](x) & \text{in } \mathbb{R}^3 \times (0, T) \\ (ii) & \partial_t m - \text{div}(m \partial_p H^\varepsilon(x, Du)) = 0 & \text{in } \mathbb{R}^3 \times (0, T) \\ (iii) & m(x, 0) = m_0(x), u(x, T) = G[m(T)](x) & \text{on } \mathbb{R}^3, \end{cases} \quad (2.7)$$

where

$$H^\varepsilon(x, p) = \frac{1}{2} |p B^\varepsilon(x)|^2 \quad \text{with} \quad B^\varepsilon := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -x_2 & x_1 & \varepsilon \end{pmatrix}. \quad (2.8)$$

Hence, explicitly, the Hamiltonian and the drift are respectively

$$H^\varepsilon(x, p) = \frac{1}{2}((p_1 - x_2 p_3)^2 + (p_2 + x_1 p_3)^2 + (\varepsilon p_3)^2)$$

$$\partial_p H^\varepsilon(x, p) = p B^\varepsilon(x)(B^\varepsilon(x))^T = (p_1 - x_2 p_3, p_2 + x_1 p_3, x_1 p_2 - x_2 p_1 + (x_1^2 + x_2^2 + \varepsilon^2) p_3)$$

while the dynamics of the generic player at point x at time t becomes

$$x'_1(s) = \alpha_1(s), \quad x'_2(s) = \alpha_2(s), \quad x'_3(s) = -x_2(s)\alpha_1(s) + x_1(s)\alpha_2(s) + \varepsilon\alpha_3(s) \tag{2.9}$$

where the control $\alpha = (\alpha_1, \alpha_2, \alpha_3)$ is chosen in $L^2([t, T]; \mathbb{R}^3)$ for minimizing the cost (1.6).

We can obtain existence, representation formula and suitable estimates of a solution to problem (2.7) which will allow us to prove Theorem 2.1 letting $\varepsilon \rightarrow 0^+$. These properties are stated in the following Proposition whose proof, together with the proof of Theorem 2.1, is postponed in section 5.

Proposition 2.1. *For any fixed $\varepsilon \in (0, 1]$ there exists a solution $(u_\varepsilon, m_\varepsilon)$ of the system (2.7) such that*

$$\int_{\mathbb{R}^3} \phi(x) dm_\varepsilon(t) = \int_{\mathbb{R}^3} \phi(\gamma_{\varepsilon,x}(t)) m_0(x) dx \quad \forall \phi \in C^0_0(\mathbb{R}^3), \forall t \in [0, T] \tag{2.10}$$

where, for a.e. $x \in \mathbb{R}^3$, $\gamma_{\varepsilon,x}$ is the unique solution to

$$x'(s) = -Du_\varepsilon(x(s), s)B^\varepsilon(x(s))(B^\varepsilon(x(s)))^T, \quad x(0) = x. \tag{2.11}$$

Moreover, there exists a positive constant C (independent of ε) such that

- a) $\|u_\varepsilon\|_\infty \leq C, \|Du_\varepsilon\|_\infty \leq C, |\partial_t u_\varepsilon(t, x)| \leq C(1 + |x_1|^2 + |x_2|^2), D^2 u_\varepsilon \leq C,$
- b) $\|m_\varepsilon\|_\infty \leq C, \mathbf{d}_1(m_\varepsilon(t_1), m_\varepsilon(t_2)) \leq C|t_2 - t_1|^{1/2}, \int_{\mathbb{R}^3} |x|^2 m_\varepsilon(x, t) dx \leq C.$

3. Formulation of the optimal control problem

In this section, we study equation (2.7)-(i) and the corresponding optimal control problem for any $\varepsilon \in [0, 1]$ (note that here we also cope the case $\varepsilon = 0$). We shall show that the value function of this control problem solves (2.7)-(i), is Lipschitz continuous and semiconcave in x . Moreover, for $\varepsilon \neq 0$, we establish an optimal synthesis result exploiting the fact that the matrix $B^\varepsilon(B^\varepsilon)^T$ is invertible. Throughout this section, since ε is a fixed parameter, we omit it when it is not essential and we assume the following hypothesis

Hypothesis 3.1. $f \in C^0([0, T], C^2(\mathbb{R}^3)), g \in C^2(\mathbb{R}^3)$ and there exists a constant C such that

$$\|f(\cdot, t)\|_{C^2(\mathbb{R}^3)} + \|g\|_{C^2(\mathbb{R}^3)} \leq C, \quad \forall t \in [0, T].$$

Definition 3.1. We consider the following optimal control problem

$$\text{minimize } J_{x,t}(\alpha) := \int_t^T \frac{1}{2} |\alpha(s)|^2 + f(x(s), s) ds + g(x(T)) \tag{3.1}$$

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

$$\mathcal{A}_\varepsilon(x, t) := \left\{ (x(\cdot), \alpha(\cdot)) \in AC([t, T]; \mathbb{R}^3) \times L^2([t, T]; \mathbb{R}^3) : (2.9) \text{ holds a.e. with } x(t) = x \right\}. \tag{3.2}$$

A couple $(x(\cdot), \alpha(\cdot)) \in \mathcal{A}_\varepsilon(x, t)$ is said to be admissible. We say that $x^*(\cdot)$ is an optimal trajectory if there is a control $\alpha^*(\cdot)$ such that $(x^*(\cdot), \alpha^*(\cdot)) \in \mathcal{A}_\varepsilon(x, t)$ is optimal for the optimal control problem in (3.1); in particular, since ε is fixed, we do not write explicitly that (x^*, α^*) depends on ε . Also, we shall refer to the system (2.9) as to the dynamics of the optimal control problem in (3.1).

Remark 3.1. Hypothesis 3.1 ensures that, for any $(x, t) \in \mathbb{R}^3 \times (0, T)$, the optimal control problem in Definition 3.1 admits a solution $(x^*(\cdot), \alpha^*(\cdot))$ thanks to the LSC with respect to the weak L^2 -topology. Moreover, just testing $J_{x,t}(\alpha^*)$ against $J_{x,t}(0)$, we get

$$\|\alpha^*\|_{L^2(t,T)} \leq C_1 := C[(T - t) + 1], \tag{3.3}$$

where C is the constant introduced in Hypothesis 3.1 (note that C_1 is independent of ε).

In particular, by Hölder inequality,

$$x^* \in C^{1/2}([t, T], \mathbb{R}^3), \tag{3.4}$$

with an 1/2-Hölder constant of the type $C(1 + |x_1| + |x_2|)$, with C independent of ε .

Definition 3.2. The value function for the cost $J_{x,t}$ defined in (3.1) is

$$u_\varepsilon(x, t) := \inf \{ J_{x,t}(\alpha) : (x(\cdot), \alpha(\cdot)) \in \mathcal{A}_\varepsilon(x, t) \}. \tag{3.5}$$

An optimal couple $(x^*(\cdot), \alpha^*)$ for the control problem in Definition 3.1 is also said to be optimal for $u(x, t)$.

The following proposition ensures that we can restrict our study on locally uniformly bounded controls.

Proposition 3.1. Let u_ε be the value function introduced in (3.5). Then, there exists a constant C_2 (depending only on T and on the constant C of Hypothesis 3.1) such that there holds

$$u_\varepsilon(x, t) = \inf \{ J_{x,t}(\alpha) : (x(\cdot), \alpha) \in \mathcal{A}_\varepsilon(x, t), \|\alpha\|_\infty \leq C_2(1 + |x_1| + |x_2|) \} \tag{3.6}$$

for any $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ and $t \in [0, T]$.

Proof. The idea of the proof is borrowed from [4, Theorem 2.1]. For $x = (x_1, x_2, x_3) \in \mathbb{R}^3$ and $t \in [0, T]$, let α be an optimal control for $u(x, t)$. For $\mu > 0$, let $I_\mu := \{s \in (t, T) : |\alpha(s)| > \mu\}$. Define

$$\alpha^\mu(s) = \begin{cases} \alpha(s) & \text{if } |\alpha(s)| \leq \mu, \\ 0 & \text{if } |\alpha(s)| > \mu. \end{cases} \tag{3.7}$$

Let $x^\mu(s)$ be the trajectory starting from x associated to the control $\alpha^\mu(s)$. We claim that

$$|x^\mu(s) - x(s)| \leq K(1 + |x_1| + |x_2| + \varepsilon) \int_{I_\mu} |\alpha(\tau)| d\tau \quad \forall s \in [t, T] \tag{3.8}$$

where K is a constant depending only on C_1 (see (3.3)) and T . Actually, for the first two components of $x^\mu(s) - x(s)$ we have

$$|x_i^\mu(s) - x_i(s)| \leq \int_t^s |\alpha_i^\mu(\tau) - \alpha_i(\tau)| d\tau = \int_{I_\mu} |\alpha_i(\tau)| d\tau \quad \forall s \in [t, T], \quad i = 1, 2. \tag{3.9}$$

For the third component, there holds

$$\begin{aligned} x_3^\mu(s) - x_3(s) &= \int_t^s [-x_2^\mu(\tau)\alpha_1^\mu(\tau) + x_1^\mu(\tau)\alpha_2^\mu(\tau) + x_2(\tau)\alpha_1(\tau) - x_1(\tau)\alpha_2(\tau) \\ &\quad + \varepsilon(\alpha_3^\mu(\tau) - \alpha_3(\tau))] d\tau \\ &= \int_t^s [(x_2(\tau) - x_2^\mu(\tau))\alpha_1^\mu(\tau) + x_2(\tau)(\alpha_1(\tau) - \alpha_1^\mu(\tau)) \\ &\quad + (x_1^\mu(\tau) - x_1(\tau))\alpha_2^\mu(\tau) + x_1(\tau)(\alpha_2^\mu(\tau) - \alpha_2(\tau)) + \varepsilon(\alpha_3^\mu(\tau) - \alpha_3(\tau))] d\tau. \end{aligned}$$

Hence from (3.4) and (3.9), we infer

$$\begin{aligned} |x_3^\mu(s) - x_3(s)| &\leq \int_{I_\mu} |\alpha_2(\tau)| d\tau \int_t^s |\alpha_1^\mu(\tau)| d\tau + [|x_2| + C_1(T - t)^{1/2}] \int_{I_\mu} |\alpha_1(\tau)| d\tau \\ &\quad + \int_{I_\mu} |\alpha_1(\tau)| d\tau \int_t^s |\alpha_2^\mu(\tau)| d\tau + [|x_1| + C_1(T - t)^{1/2}] \int_{I_\mu} |\alpha_2(\tau)| d\tau \\ &\quad + \varepsilon \int_{I_\mu} |\alpha_3(\tau)| d\tau. \end{aligned}$$

Moreover, by Hölder inequality and (3.3), we have

$$\int_t^s |\alpha_i^\mu(\tau)| d\tau \leq \sqrt{s-t} \|\alpha\|_2 \leq C_1 \sqrt{T-t}, \quad i = 1, 2.$$

Replacing the last inequality in the previous one, we accomplish the proof of the claim (3.8).

Now, the definition of the cost $J_{x,t}(\alpha)$ in (3.1) and the Lipschitz continuity of f and g yield

$$\begin{aligned} J_{x,t}(\alpha^\mu) - J_{x,t}(\alpha) &= \\ &= \int_t^T \frac{1}{2} |\alpha^\mu(s)|^2 + f(x^\mu(s), s) ds + g(x^\mu(T)) - \int_t^T \frac{1}{2} |\alpha(s)|^2 + f(x(s), s) ds - g(x(T)) \\ &\leq - \int_{I_\mu} \frac{1}{2} |\alpha(s)|^2 ds + L_f \int_t^T |x^\mu(s) - x(s)| ds + L_g |x^\mu(T) - x(T)| \\ &\leq \int_{I_\mu} \left(-\frac{1}{2} |\alpha(s)|^2 + K(L_f(T-t) + L_g)(1 + |x_1| + |x_2| + \varepsilon) |\alpha(s)| \right) ds, \end{aligned}$$

where the last inequality comes from (3.8). Hence, if I_μ has positive measure, for

$$\mu > 2K(L_f(T-t) + L_g)(1 + |x_1| + |x_2| + \varepsilon),$$

the last integrand is negative for every $s \in I_\mu$ which contradicts the optimality of α . This implies that these I_μ have null measure and, in particular,

$$\|\alpha\|_\infty \leq 2K(L_f(T-t) + L_g)(1 + |x_1| + |x_2| + \varepsilon).$$

Since the choice of μ is independent of $\varepsilon \in [0, 1]$, we get the result. \square

3.1. Necessary conditions and regularity for the optimal trajectories

The application of the Maximum Principle (see [19, Theorem 22.17]) yields the following necessary conditions.

Proposition 3.2. *Let (x^*, α^*) be optimal for the optimal control problem in (3.1). Then, there exists an arc $p \in AC([t, T]; \mathbb{R}^3)$, hereafter called the costate, such that*

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

$$\left\{ \begin{aligned} x'_1 &= p_1 - x_2 p_3 \\ x'_2 &= p_2 + x_1 p_3 \\ x'_3 &= (x_1^2 + x_2^2 + \varepsilon^2) p_3 + x_1 p_2 - x_2 p_1 \\ p'_1 &= -(p_2 + x_1 p_3) p_3 + f_{x_1}(x, s) \\ p'_2 &= (p_1 - x_2 p_3) p_3 + f_{x_2}(x, s) \\ p'_3 &= f_{x_3}(x, s) \end{aligned} \right. \tag{3.10}$$

with the mixed boundary conditions

$$x(t) = x, \quad p(T) = -Dg(x(T)). \tag{3.11}$$

2. The optimal control α^* verifies for a.e. $s \in [t, T]$

$$\alpha_1(s) = p_1 - x_2 p_3, \quad \alpha_2(s) = p_2 + x_1 p_3, \quad \alpha_3(s) = \varepsilon p_3. \tag{3.12}$$

Remark 3.2. Let us observe that equations (3.10) and (3.12) can be rewritten in terms of the vector fields as follows

$$\begin{aligned} x'_1 &= X_1 p, & x'_2 &= X_2 p, & x'_3 &= -x_2 X_1 p + x_1 X_2 p + \varepsilon^2 p_3, \\ p'_1 &= -p_3 X_2 p + f_{x_1}(x, s), & p'_2 &= p_3 X_1 p + f_{x_2}(x, s), & p'_3 &= f_{x_3}(x, s) \end{aligned}$$

and respectively

$$\alpha_1(s) = X_1 p(s), \quad \alpha_2(s) = X_2 p(s), \quad \alpha_3(s) = \varepsilon p_3(s).$$

Corollary 3.1. Let (x^*, α^*) be optimal for the optimal control problem in (3.1). Then:

1. The unique solution of the Cauchy problem

$$\begin{cases} \pi'_1 = -(\pi_2 + x_1^* \pi_3) \pi_3 + f_{x_1}(x^*, s), \\ \pi'_2 = (\pi_1 - x_2^* \pi_3) \pi_3 + f_{x_2}(x^*, s), \\ \pi'_3 = f_{x_3}(x^*, s), \\ \pi(T) = -Dg(x^*(T)), \end{cases}$$

is the costate p associated to (x^*, α^*) as in Proposition 3.2.

2. The optimal α^* is a feedback control and it is uniquely expressed by

$$\alpha_1^*(s) = p_1 - x_2^* p_3, \quad \alpha_2^*(s) = p_2 + x_1^* p_3, \quad \alpha_3^*(s) = \varepsilon p_3$$

where p is the costate associated to (x^*, α^*) .

3. The functions x^* and α^* are of class C^1 . In particular equations (3.10) and (3.12) hold for every $s \in [t, T]$.

4. Assume that, for some $k \in \mathbb{N}$, $D_x f \in C^k$. Then, the costate p and the control α^* are of class C^{k+1} and x^* is of class C^{k+2} .

Proof. Point 1 is an immediate consequence of (3.10) together with the endpoint condition (3.11). Point 2 follows from (3.12).

3. Since x^* and p are continuous, the continuity of α^* follows from (3.12). The dynamics (2.9) yield that $x^* \in C^1$. Relations (3.10) and (3.12) imply that p and α^* are of class C^1 . By a standard bootstrap inductive argument still using (2.9), (3.10) and (3.12), we get point 4. \square

Proposition 3.3. For $\varepsilon \neq 0$, the optimal trajectories are unique after the initial time: if $x^*(\cdot)$ is an optimal trajectory for $u_\varepsilon(x, t)$, then for every $t < \tau < T$ there are no other optimal trajectories for $u_\varepsilon(x^*(\tau), \tau)$ other than $x^*(\cdot)$ restricted to $[\tau, T]$.

Proof. Let y^* be an optimal trajectory for $u(x^*(\tau), \tau)$; the concatenation z^* of x^* with y^* at τ is still optimal for $u_\varepsilon(x, t)$. Let p and q be respectively the costate of x^* and of z^* . By point (4) both x^* and z^* are C^1 . Since the matrix $B^\varepsilon(x)(B^\varepsilon(x))^T$ is invertible, we denote by $\beta(x)$ its inverse and, from the first three lines in (3.10), we get

$$p(s) = \beta(x^*(s))\dot{x}^*(s) \quad \text{and} \quad q(s) = \beta(z^*(s))\dot{z}^*(s) \quad \forall s \in (t, T).$$

Since $x^*(\cdot) = z^*(\cdot)$ in $[t, \tau]$, we get: $p(\tau) = q(\tau)$. In conclusion, both (x^*, p) and (z^*, q) solve the same Cauchy problem (3.10) on $(\tau, T]$ with the same data at time τ . The Cauchy-Lipschitz theorem ensures that they coincide. \square

3.2. 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 (2.7)-(i) with m fixed, namely

$$\begin{cases} -\partial_t u_\varepsilon + H^\varepsilon(x, Du_\varepsilon) = f(x, t) & \text{in } \mathbb{R}^3 \times (0, T), \\ u_\varepsilon(x, T) = g(x) & \text{on } \mathbb{R}^3 \end{cases} \quad (3.13)$$

where H^ε is defined in (2.8). Under Hypothesis 3.1, we shall prove Lipschitz continuity and semiconcavity of u_ε . As a first step, in the next lemma we show that the solution u_ε of (3.13) can be represented as the value function of the control problem defined in (3.5).

Lemma 3.1. *Under Hypothesis 3.1, the value function u_ε defined in (3.5) is the unique bounded viscosity solution to problem (3.13) and there exists a constant C independent of ε such that*

$$\|u_\varepsilon\|_\infty \leq C. \quad (3.14)$$

Proof. The proof comes from classical results: see for instance [5, Proposition III.3.5] and [4, Theorem 3.1]). The bound of u_ε uniform on ε is obtained taking as admissible control $\alpha = 0$ in (3.5). \square

Lemma 3.2. *The value function u_ε fulfills the following properties*

1. u_ε is Lipschitz continuous with respect to the spatial variable x uniformly on ε ,
2. u_ε is locally Lipschitz continuous with respect to the time variable t with a Lipschitz constant $C(1 + |x_1|^2 + |x_2|^2)$ where C is a constant independent of ε .

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 . For simplicity we write “ u ” instead of “ u_ε ”.

1. Let t be fixed. We follow the proof of [16, Lemma 4.7]. From Remark 3.1 we know that there exists $\alpha(\cdot)$ optimal control for $u(x, t)$ and $x(\cdot)$ optimal trajectory i.e.:

$$u(x_1, x_2, x_3, t) = \int_t^T \frac{1}{2} |\alpha(s)|^2 + f(x(s), s) ds + g(x(T)). \quad (3.15)$$

We consider the path $x^*(s)$ starting from $y = (y_1, y_2, y_3)$, with control α . We have

$$\begin{aligned}
 x_1^*(s) &= y_1 + \int_t^s \alpha_1(\tau) d\tau = y_1 - x_1 + x_1(s) \\
 x_2^*(s) &= y_2 + \int_t^s \alpha_2(\tau) d\tau = y_2 - x_2 + x_2(s) \\
 x_3^*(s) &= y_3 - \int_t^s \alpha_1(\tau)x_2^*(\tau) d\tau + \int_t^s \alpha_2(\tau)x_1^*(\tau) d\tau + \varepsilon \int_t^s \alpha_3(\tau) d\tau \\
 &= y_3 - (y_2 - x_2) \int_t^s \alpha_1(\tau) d\tau + (y_1 - x_1) \int_t^s \alpha_2(\tau) d\tau + \varepsilon \int_t^s \alpha_3(\tau) d\tau \\
 &\quad + \int_t^s (-\alpha_1(\tau)x_2(\tau) + \alpha_2(\tau)x_1(\tau)) d\tau \\
 &= x_3(s) + (y_3 - x_3) - (y_2 - x_2) \int_t^s \alpha_1(\tau) d\tau + (y_1 - x_1) \int_t^s \alpha_2(\tau) d\tau + \varepsilon \int_t^s \alpha_3(\tau) d\tau \\
 &= x_3(s) + (y_3 - x_3) - (y_2 - x_2) \int_t^s \alpha_1(\tau) d\tau + (y_1 - x_1) \int_t^s \alpha_2(\tau) d\tau.
 \end{aligned}$$

Using the Lipschitz continuity of f we get

$$\begin{aligned}
 f(x^*(s), s) &\leq f(x(s), s) + L(|y_1 - x_1| + |y_2 - x_2| + |y_3 - x_3| + \\
 &\quad + |y_2 - x_2|\sqrt{s-t}\|\alpha_1\|_2 + |y_1 - x_1|\sqrt{s-t}\|\alpha_2\|_2)
 \end{aligned}$$

and from the L^2 uniform estimate for α_1 and α_2 in (3.3) we get

$$f(x^*(s), s) - f(x(s), s) \leq C_T(|y_1 - x_1| + |y_2 - x_2| + |y_3 - x_3|).$$

By the same calculations for g and substituting equality (3.15) in

$$u(y_1, y_2, y_3, t) \leq \int_t^T \frac{1}{2} |\alpha(s)|^2 + f(x^*(s), s) ds + g(x^*(T)),$$

we get

$$u(y_1, y_2, y_3, t) \leq u(x_1, x_2, x_3, t) + C_T(|y_1 - x_1| + |y_2 - x_2| + |y_3 - x_3|).$$

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

2. For proving the Lipschitz continuity w.r.t. t , we note that

$$\begin{aligned} |x(s) - x| &\leq C_T(s - t)(\|\alpha_1\|_\infty|x_2| + \|\alpha_2\|_\infty|x_1| + \varepsilon\|\alpha_3\|_\infty) \\ &\leq C_T(1 + |x_1|^2 + |x_2|^2 + \varepsilon)(s - t) \\ &\leq C_T(1 + |x_1|^2 + |x_2|^2), \end{aligned}$$

where the second inequality is due to the L^∞ -bound for optimal controls established in Proposition 3.1. We accomplish the proof following the same arguments as those in the proof of [16, Lemma 4.7]. \square

In the following lemma we establish the semiconcavity of u w.r.t. x ; we recall here below the definition of semiconcavity with linear modulus and we refer the reader to the monograph [14] for further properties.

Definition 3.3. Let $u : \mathbb{R}^d \rightarrow \mathbb{R}$. We say that u is *semiconcave* (with linear modulus) if there exists a constant $C \geq 0$ such that for all $\lambda \in [0, 1]$,

$$\lambda u(y) + (1 - \lambda)u(x) - u(\lambda y + (1 - \lambda)x) \leq C\lambda(1 - \lambda)|y - x|^2 \quad \forall x, y \in \mathbb{R}^d.$$

Lemma 3.3. Under Hypothesis 3.1, the value function u_ε , defined in (3.5), is semiconcave w.r.t. x with a linear modulus independent on ε .

Proof. For any $x, y \in \mathbb{R}^3$ 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 respectively the corresponding optimal trajectory for $u(x_\lambda, t)$; for $s \in [t, T]$ there hold

$$\begin{aligned} x_{\lambda,i}(s) &= x_{\lambda,i} + \int_t^s \alpha_i(\tau) d\tau, \quad \text{for } i = 1, 2 \\ x_{\lambda,3}(s) &= x_{\lambda,3} - \int_t^s \alpha_1(\tau)x_{\lambda,2}(\tau) d\tau + \int_t^s \alpha_2(\tau)x_{\lambda,1}(\tau) d\tau + \varepsilon \int_t^s \alpha_3(\tau) d\tau. \end{aligned}$$

Let $x(s)$ and $y(s)$ satisfy (1.5) 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 [16, 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 third component $x_3(s)$ since the calculations for $x_1(s)$ and $x_2(s)$ are the same as in [16]. We have

$$x_3(s) = x_3 - \int_t^s \alpha_1(\tau)x_2(\tau) d\tau + \int_t^s \alpha_2(\tau)x_1(\tau) d\tau + \varepsilon \int_t^s \alpha_3(\tau) d\tau$$

$$= x_3 - x_{\lambda,3} + x_{\lambda,3}(s) - \int_t^s \alpha_1(\tau)(x_2(\tau) - x_{\lambda,2}(\tau)) d\tau + \int_t^s \alpha_2(\tau)(x_1(\tau) - x_{\lambda,1}(\tau)) d\tau.$$

Since $x_3 - x_{\lambda,3} = (1 - \lambda)(x_3 - y_3)$ and

$$x_i(\tau) - x_{\lambda,i}(\tau) = (1 - \lambda)(x_i - y_i) \quad \text{for } i = 1, 2, \tag{3.16}$$

we get

$$x_3(s) - x_{\lambda,3}(s) = (1 - \lambda) \left[x_3 - y_3 - (x_2 - y_2) \int_t^s \alpha_1(\tau) d\tau + (x_1 - y_1) \int_t^s \alpha_2(\tau) d\tau \right]. \tag{3.17}$$

Analogously for $y(s)$: since $y_3 - x_{\lambda,3} = \lambda(y_3 - x_3)$ and

$$y_i(\tau) - x_{\lambda,i}(\tau) = \lambda(y_i - x_i) \quad \text{for } i = 1, 2, \tag{3.18}$$

we get

$$y_3(s) - x_{\lambda,3}(s) = \lambda \left[(y_3 - x_3) + (x_2 - y_2) \int_t^s \alpha_1(\tau) d\tau - (x_1 - y_1) \int_t^s \alpha_2(\tau) d\tau \right]. \tag{3.19}$$

For the sake of brevity we provide the explicit calculations only for f omitting the analogous ones for g ; and we write $f(x_1, x_2, x_3) := f(x_1, x_2, x_3, s)$. We have

$$\begin{aligned} &\lambda f(x(s)) + (1 - \lambda) f(y(s)) = \\ &\lambda f(x_1(s), x_2(s), x_{\lambda,3}(s) + (1 - \lambda)(x_3 - y_3 - (x_2 - y_2) \int_t^s \alpha_1(\tau) d\tau + (x_1 - y_1) \int_t^s \alpha_2(\tau) d\tau)) + \\ &+ (1 - \lambda) f(y_1(s), y_2(s), x_{\lambda,3}(s) + \lambda(y_3 - x_3 + (x_2 - y_2) \int_t^s \alpha_1(\tau) d\tau - (x_1 - y_1) \int_t^s \alpha_2(\tau) d\tau)). \end{aligned}$$

Since for $i = 1, 2$ there holds

$$\lambda \partial_{x_i} f(x_{\lambda}(s))(x_i(s) - x_{\lambda,i}(s)) + (1 - \lambda) \partial_{x_i} f(x_{\lambda}(s))(y_i(s) - x_{\lambda,i}(s)) = 0,$$

the Taylor expansion of f centered in $x_{\lambda}(s)$ gives:

$$\begin{aligned} \lambda f(x(s)) + (1 - \lambda) f(y(s)) &= \lambda(f(x_{\lambda}(s)) + Df(x_{\lambda}(s))(x(s) - x_{\lambda}(s)) + R_1) \\ &\quad + (1 - \lambda)(f(x_{\lambda}(s)) + Df(x_{\lambda}(s))(y(s) - x_{\lambda}(s)) + R_2) \\ &= \lambda \left(f(x_{\lambda}(s)) + \partial_{x_3} f(x_{\lambda}(s))(1 - \lambda)(x_3 - y_3 - (x_2 - y_2) \int_t^s \alpha_1(\tau) d\tau \right. \\ &\quad \left. + (x_1 - y_1) \int_t^s \alpha_2(\tau) d\tau) + R_1 \right) \end{aligned}$$

$$\begin{aligned}
 &+ (1 - \lambda) \left(f(x_\lambda(s)) + \partial_{x_3} f(x_\lambda(s)) \lambda (y_3 - x_3 + (x_2 - y_2) \int_t^s \alpha_1(\tau) d\tau) \right) \\
 &\quad - (1 - \lambda) \left((x_1 - y_1) \int_t^s \alpha_2(\tau) d\tau + R_2 \right) \\
 &= f(x_\lambda(s)) + \lambda R_1 + (1 - \lambda) R_2,
 \end{aligned}$$

where R_1 and R_2 are the error terms of the expansion, namely

$$\begin{aligned}
 \lambda R_1 + (1 - \lambda) R_2 &= \frac{1}{2} \lambda (x(s) - x_\lambda(s)) D^2 f(\xi_1) (x(s) - x_\lambda(s))^T \\
 &\quad + \frac{1}{2} (1 - \lambda) (y(s) - x_\lambda(s)) D^2 f(\xi_2) (y(s) - x_\lambda(s))^T,
 \end{aligned}$$

for suitable $\xi_1, \xi_2 \in \mathbb{R}^3$.

Using relations (3.16)-(3.19) and the L^2 uniform estimate of α in (3.3) (which is independent of ε), we obtain

$$\begin{cases} |x_i(s) - x_{\lambda,i}(s)| |x_j(s) - x_{\lambda,j}(s)| \leq C(1 - \lambda)^2 |x - y|^2 & i, j = 1, 2, 3 \\ |y_i(s) - x_{\lambda,i}(s)| |y_j(s) - x_{\lambda,j}(s)| \leq C\lambda^2 |x - y|^2 & i, j = 1, 2, 3 \end{cases}$$

for some positive constant C . Then, using Hypothesis 3.1, possibly increasing C , we get

$$\lambda R_1 + (1 - \lambda) R_2 \leq C\lambda(1 - \lambda) |x - y|^2,$$

and, in particular,

$$\lambda f(x(s)) + (1 - \lambda) f(y(s)) \leq f(x_\lambda(s)) + C\lambda(1 - \lambda) |x - y|^2$$

which amounts to the semiconcavity of u . \square

Lemma 3.4 (Optimal synthesis). *Let u_ε be the unique bounded viscosity solution to (3.13) founded in Lemma 3.1.*

a) *Let $\gamma \in AC([t, T])$ be such that*

$$u_\varepsilon(\cdot, s) \text{ is differentiable at } \gamma(s) \text{ for almost every } s \in (t, T) \tag{3.20}$$

and

$$\dot{\gamma}(t) = -Du_\varepsilon(\gamma(t), t) B^\varepsilon(\gamma(t)) (B^\varepsilon(\gamma(t)))^T, \quad \gamma(0) = x. \tag{3.21}$$

Then, the control law $\alpha(s) = -Du_\varepsilon(\gamma(s), s) B^\varepsilon(\gamma(s))$ is optimal for $u_\varepsilon(x, t)$.

b) For $\varepsilon \neq 0$, if $u_\varepsilon(\cdot, t)$ is differentiable at x , then problem (3.21) has a unique solution which moreover coincides with the optimal trajectory. In particular, for a.e. x , there exists a unique optimal trajectory for $u_\varepsilon(x, 0)$.

Proof. We shall follow the same arguments as those used in [16, Lemma 4.11] (see also [2, Lemma 3.5] and [33, Lemma 3.5] for similar arguments). So we only illustrate the main novelties in the proof and we refer the reader to those papers for the details. Since ε is fixed, for simplicity we write “ u ” instead of “ u_ε ”.

(a). Let γ be a curve as in the statement; we claim that γ is bounded and Lipschitz continuous. Indeed, the differential equation (3.21) reads

$$\gamma'_1 = -u_{x_1} + \gamma_2 u_{x_3}, \quad \gamma'_2 = -u_{x_2} - \gamma_1 u_{x_3}, \quad \gamma'_3 = \gamma_2 u_{x_1} - \gamma_1 u_{x_2} - (\gamma_1^2 + \gamma_2^2 + \varepsilon^2) u_{x_3}.$$

Due to the structure of B^ε , the first two equations do not depend on γ_3 . We define $\xi := (\gamma_1, \gamma_2)$. By the Cauchy-Schwarz inequality and the Lipschitz continuity of u found in Lemma 3.2, there exists a constant C such that $|\xi'| \leq C(|\xi| + 1)$. By Gronwall Lemma, we deduce that ξ is bounded and Lipschitz continuous and, consequently, that γ_1 and γ_2 are bounded and Lipschitz continuous. Using these properties and again the Lipschitz continuity of u , in the third component of (3.21), we obtain that also γ_3 is Lipschitz continuous. Our claim is proved.

The Lipschitz continuity of γ and of u entail that the function $s \mapsto u(\gamma(s), s)$ is Lipschitz continuous. The rest of the proof follows the same arguments of the aforementioned papers and it relies on Lebourg’s mean value Theorem, Charathéodory Theorem and properties of semiconcave functions.

(b). We claim that if $Du_\varepsilon(x, t)$ exists then the set $\{\alpha(t), \alpha \text{ optimal for } u(x, t)\}$ is a singleton and that there holds $\alpha(t) = -Du_\varepsilon(x, t)B^\varepsilon(x)$. Indeed, if $\alpha(\cdot)$ is optimal for $u_\varepsilon(x, t)$ and $x(\cdot)$ is the corresponding optimal trajectory, then

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

and $x(\cdot)$ and $\alpha(\cdot)$ satisfy the necessary conditions for optimality proved in Proposition 3.2.

Take $h \in \mathbb{R}^3$ and consider the solution $y(\cdot)$ of (2.9) starting at point $y = (x_1 + h_1, x_2 + h_2, x_3 - x_2 h_1 + x_1 h_2 + \varepsilon h_3)$ at time t and obeying to the control α . Using the calculations in the proof of point 1. of Lemma 3.2 we get

$$\begin{aligned} y_1(s) &= h_1 + x_1(s), \\ y_2(s) &= h_2 + x_2(s), \\ y_3(s) &= x_3(s) + x_1 h_2 - x_2 h_1 + \varepsilon h_3 - h_2 \int_t^s \alpha_1(\tau) d\tau + h_1 \int_t^s \alpha_2(\tau) d\tau, \end{aligned}$$

and

$$u_\varepsilon(y, t) - u_\varepsilon(x, t) \leq \int_t^T (f(y(s), s) - f(x(s), s)) ds + g(y(T)) - g(x(T)). \tag{3.22}$$

From the differentiability of u w.r.t. x and the arbitrariness of the components of h we get

$$Du_\varepsilon(x, t)B^\varepsilon(x) = (I_1, I_2, I_3)$$

where

$$\begin{aligned}
 I_1 &= \int_t^T f_{x_1}(x(s), s)ds + \int_t^T f_{x_3}(x(s), s)\left(\int_t^s \alpha_2(\tau)d\tau - x_2\right)ds + g_{x_1}(x(T)) + \\
 &\quad g_{x_3}(x(T))\left(\int_t^T \alpha_2(\tau)d\tau - x_2\right), \\
 I_2 &= \int_t^T f_{x_2}(x(s), s)ds + \int_t^T f_{x_3}(x(s), s)\left(x_1 - \int_t^s \alpha_1(\tau)d\tau\right)ds + g_{x_2}(x(T)) + \\
 &\quad g_{x_3}(x(T))\left(x_1 - \int_t^T \alpha_1(\tau)d\tau\right), \\
 I_3 &= \varepsilon \int_t^T f_{x_3}(x(s), s)ds + \varepsilon g_{x_3}(x(T)).
 \end{aligned}$$

From (3.10) and (3.11) we have that $f_{x_3}(x(s), s) = p'_3(s)$ and $-g_{x_3}(x(T)) = p_3(T)$. Then, integrating by parts and using (3.12) and (3.10), we get $I_1 = -\alpha_1(t)$, $I_2 = -\alpha_2(t)$ and $I_3 = -\alpha_3(t)$, i.e. $\alpha(t) = -Du_\varepsilon(x, t)B^\varepsilon(x)$. This uniquely determines the value of $\alpha(\cdot)$ at time t .

In particular the value of $p(t)$ is uniquely determined by relations (3.12) (here, $\varepsilon \neq 0$ is needed); hence, system (3.10) becomes a system of differential equations with condition at time t and admits a unique solution by the Cauchy-Lipschitz theorem and $x(\cdot)$ is the unique optimal trajectory starting from x associated to the unique optimal control $\alpha(s)$ given by (3.12). Then from (3.10) $x(\cdot)$ is a solution of (3.21) and by point a) is the unique solution. \square

4. The continuity equation

This section is devoted to study equation (2.7)-(ii), namely, to study

$$\begin{cases} \partial_t m - \operatorname{div}(m DuB^\varepsilon(x)(B^\varepsilon(x))^T) = 0 & \text{in } \mathbb{R}^3 \times (0, T) \\ m(x, 0) = m_0(x) & \text{on } \mathbb{R}^3 \end{cases} \tag{4.1}$$

where u is the unique bounded (viscosity) solution to problem

$$\begin{cases} -\partial_t u + \frac{1}{2}|DuB^\varepsilon(x)|^2 = F[\overline{m}(t)](x) & \text{in } \mathbb{R}^3 \times (0, T) \\ u(x, T) = G[\overline{m}(T)](x) & \text{on } \mathbb{R}^3 \end{cases} \tag{4.2}$$

with $\bar{m} \in C^0([0, T], \mathcal{P}_1(\mathbb{R}^3))$ (see Lemma 3.1 for the existence, uniqueness and boundedness of u). Since now on, throughout this section, $\varepsilon \in (0, 1]$ and \bar{m} are fixed.

Now we deal with the existence and uniform estimates of the solution m of (4.1).

Theorem 4.1. *Under assumptions (H1)-(H3), for any $\bar{m} \in C^0([0, T], \mathcal{P}_1(\mathbb{R}^3))$, problem (4.1) has a solution m in the sense of Definition 2.1. The function m belongs to $C^{1/2}([0, T], \mathcal{P}_1(\mathbb{R}^3)) \cap L^\infty(0, T; \mathcal{P}_2(\mathbb{R}^3))$ with a density in $L^\infty(\mathbb{R}^3 \times (0, T))$. Moreover, m fulfills (2.10)-(2.11) and there exists a constant K independent of \bar{m} and of ε such that*

$$\|m\|_\infty \leq K, \quad \mathbf{d}_1(m(t_1), m(t_2)) \leq K|t_2 - t_1|^{1/2}, \quad \int_{\mathbb{R}^3} |x|^2 m(x, t) dx \leq K. \quad (4.3)$$

The proof of this theorem is postponed at the end of this section. It borrows some arguments of the proof of [33, Proposition 3.1] (see also [17, Theorem 5.1] and [16, Theorem 4.20]). More precisely, we shall use a vanishing viscosity approach with the Euclidean Laplacian in the whole system and a truncation argument only in the continuity equation. The vanishing viscosity approach permits to exploit the well posedness of uniformly parabolic equations while the truncation argument permits to overcome the issue of coefficients which grow “too much” as $x \rightarrow \infty$.

For $\sigma \in (0, 1]$, we consider the problem

$$\begin{cases} (i) & -\partial_t u - \sigma \Delta u + H^\varepsilon(x, Du) = F[\bar{m}(t)](x) & \text{in } \mathbb{R}^3 \times (0, T) \\ (ii) & \partial_t m - \sigma \Delta m - \operatorname{div}(m Du B^{\varepsilon, N}(x)(B^{\varepsilon, N}(x))^T) = 0 & \text{in } \mathbb{R}^3 \times (0, T) \\ (iii) & m(x, 0) = m_0(x), u(x, T) = G[\bar{m}(T)](x) & \text{on } \mathbb{R}^3, \end{cases} \quad (4.4)$$

where

$$B^{\varepsilon, N}(x) := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\psi_N(x_2) & \psi_N(x_1) & \varepsilon \end{pmatrix}, \quad \psi_N(\xi) := \begin{cases} \xi & \text{if } |\xi| \leq N \\ 0 & \text{if } |\xi| \geq 2N \end{cases}$$

with $\psi_N \in C^2(\mathbb{R})$, $|\psi_N(\xi)| \leq |\xi|$ for any $\xi \in \mathbb{R}$, $\|\psi_N\|_{L^\infty} \leq 2N$, $\|\psi'_N\|_{L^\infty} + \|\psi''_N\|_{L^\infty} \leq K$ (for a constant K independent of N).

In order to prove Theorem 4.1, it is expedient to establish several properties of the solution (u^σ, m^σ) to system (4.4): the following lemmata collect existence, uniqueness and other properties of u^σ and respectively m^σ .

Lemma 4.1. *There exists a unique bounded classical solution u^σ to equation (4.4)-(i) with terminal condition (4.4)-(iii). Moreover there exists a positive constant C (independent of ε, σ, N and \bar{m}) such that*

- a) $\|u^\sigma\|_\infty \leq C$
- b) $\|Du^\sigma\|_\infty \leq C, |\partial_t u^\sigma(t, x)| \leq C(1 + |x_1|^2 + |x_2|^2)$
- c) $D^2 u^\sigma \leq C.$

Proof. It is enough to invoke classical results on parabolic equations for the existence of a solution and the comparison principle (see for instance, [9,32] and also [21]) using super- and subsolutions of the form $w^\pm = \pm C(T + 1 - t)$. In particular, we get point (a).

Moreover, since the coefficients of B^ε have linear growth at infinity, the bounded solution u^σ of (4.4)-(i) is the value function of a stochastic optimal control problem, namely

$$u_\sigma(x, t) = \min \mathbb{E} \left(\int_t^T \left[\frac{1}{2} |\alpha(\tau)|^2 + F[\bar{m}_\tau](Y_\tau) \right] d\tau + G[\bar{m}_T](Y_T) \right) \tag{4.5}$$

where, in $[t, T]$, $Y(\cdot)$ obeys to a stochastic differential equation

$$dY = \alpha(t)B^\varepsilon(Y_t)^T dt + \sqrt{2\sigma}dW_t, \quad Y_t = x \tag{4.6}$$

and W_t is a standard 3-dimensional Brownian motion. Following the procedure used in Lemma 3.2, point (1), for the deterministic case, we can prove the locally uniform Lipschitz continuity of u^σ . Actually when evaluating the difference between two random variables obeying to (4.6) with the same control, the contribution of the diffusion term is null. Hence we get the first bound of (b). Similarly, still using formula (4.5) and the procedure used in Lemma 3.2, point (2), we get the second bound of (b); using a procedure as in Lemma 3.3 we get the uniform semiconcavity of u^σ , namely point (c). \square

Lemma 4.2. *There exists a unique bounded classical solution $m_{\sigma,N}$ to problem (4.4)-(ii) with initial condition (4.4)-(iii). Moreover, there exists a constant C_N (independent of ε, σ and \bar{m}) such that $0 < m_{\sigma,N} \leq C_N$ in $\mathbb{R}^3 \times (0, T)$.*

Proof. Since σ is fixed, for simplicity we write “ u ” instead of “ u_σ ”. By the regularity of u (see Lemma 4.1), the equation (4.4)-(ii) can be written as

$$\partial_t m - \sigma \Delta m - Dm \cdot (DuB^{\varepsilon,N}(B^{\varepsilon,N})^T) - m \operatorname{div}(DuB^{\varepsilon,N}(B^{\varepsilon,N})^T) = 0. \tag{4.7}$$

By Lemma 4.1-(b), we have $\|DuB^{\varepsilon,N}(B^{\varepsilon,N})^T\|_\infty \leq CN^2$ and also, using standard estimate on the heat kernel, $\|D^2u\|_\infty \leq C$ (where C depends on σ). By standard results on parabolic equations, we infer existence and uniqueness of a bounded solution $m_{\sigma,N}$ to (4.4). From the assumptions on m_0 and Harnack inequality (see for example [28, Theorem 2.1, p. 13]) we get that $m_{\sigma,N}(\cdot, t) > 0$ for $t > 0$.

Let us now prove that $m_{\sigma,N}$ fulfills an upper bound, which is independent of ε, σ and \bar{m} . To this end, we assume for the moment that there exists a positive constant k_N (independent of ε, σ and \bar{m}) such that

$$\operatorname{div}(DuB^{\varepsilon,N}(B^{\varepsilon,N})^T) \leq k_N. \tag{4.8}$$

By estimate (4.8), the positivity of $m_{\sigma,N}$ and assumption (2.2), from equation (4.7) we get

$$\partial_t m - \sigma \Delta m - Dm \cdot (DuB^{\varepsilon,N}(B^{\varepsilon,N})^T) - mk_N \leq 0, \quad m(x, 0) \leq C.$$

Then, using the comparison principle with a supersolution of the form $w = C'e^{C't}$ (with C' depending only on k_N), we obtain that $m_{\sigma,N} \leq C_N$.

It remains to prove (4.8). We denote by I the left hand side. There holds

$$\begin{aligned}
 I &= \partial_{11}u - 2\psi_N(x_2)\partial_{13}u + \partial_{22}u + 2\psi_N(x_1)\partial_{23}u + (\psi_N(x_2)^2 + \psi_N(x_1)^2 + \varepsilon^2)\partial_{33}u \\
 &= \xi_1 D^2 u \xi_1^T + \xi_2 D^2 u \xi_2^T + \xi_3 D^2 u \xi_3^T
 \end{aligned}$$

where $\xi_1 = (1, 0, -\psi_N(x_2))$, $\xi_2 = (0, 1, \psi_N(x_1))$ and $\xi_3 = (0, 0, \varepsilon)$. By Lemma 4.1-(c), we obtain

$$I \leq C(2 + 2\|\psi_N\|_\infty^2 + \varepsilon^2) \leq C(3 + 8N^2)$$

where the last inequality is due to our assumption on ψ_N . Choosing $k_N = C(3 + 8N^2)$ we accomplish the proof of (4.8). □

Lemma 4.3. *There exists a constant K_N , depending on N (but independent of ε, σ and \bar{m}) such that the function $m_{\sigma,N}$ found in Lemma 4.2 verifies*

- a) $\mathbf{d}_1(m_{\sigma,N}(t_1), m_{\sigma,N}(t_2)) \leq K_N(t_2 - t_1)^{1/2}$ for any $0 \leq t_1 < t_2 \leq T$
- b) $\int_{\mathbb{R}^3} |x|^2 m_{\sigma,N}(x, t) dx \leq K_N(\int_{\mathbb{R}^3} |x|^2 dm_0(x) + 1)$.

Proof. We take account that $m_{\sigma,N}$ can be interpreted as the law of the following stochastic process

$$dY_t = -Du(Y_t, t)B^{\varepsilon,N}(Y_t)(B^{\varepsilon,N}(Y_t))^T dt + \sqrt{2\sigma}dW_t, \quad Y_0 = Z_0, \tag{4.9}$$

where $\mathcal{L}(Z_0) = m_0$.

By standard arguments, setting

$$m_{\sigma,N}(t) := \mathcal{L}(Y_t), \tag{4.10}$$

we know that $m_{\sigma,N}(t)$ is absolutely continuous with respect to Lebesgue measure, and that if $m_{\sigma,N}(\cdot, t)$ is the density of $m_{\sigma,N}(t)$, then $m_{\sigma,N}$ is the weak solution to (4.4)-(ii) with $m_{\sigma,N}|_{t=0} = m_0$ (from Ito's Theorem, since the drift is bounded, Proposition 3.6 Chapter 5 [27], p. 303, and the book [10]). Here we have used the bound on $\|Du^\sigma\|$ given in Lemma 4.1.

b) Noting that

$$\int_{\mathbb{R}^3} |x|^2 dm_{\sigma,N}(t)(x) = \mathbb{E}(|Y_t|^2),$$

the desired estimate can be obtained applying [27, Estimate 3.17, p. 306] (see also p. 389).

a) For $t_2 \geq t_1$, it is well known (for instance, see [16, Lemma 3.4 (proof)]) that

$$\mathbf{d}_1(m_{\sigma,N}(t_1), m_{\sigma,N}(t_2)) \leq \mathbb{E}(|Y_{t_1} - Y_{t_2}|).$$

Using this inequality and the boundedness of the drift term in (4.9) with a constant C_N , we get that there exists a constant K_N such that

$$\mathbb{E}(|Y_{t_1} - Y_{t_2}|) \leq \mathbb{E} \left(\int_{t_1}^{t_2} C_N d\tau + \sqrt{2\sigma} |B_{t_2} - B_{t_1}| \right) \leq K_N \sqrt{t_2 - t_1},$$

where we have used estimate [27, (3.17) p. 306] to the term $|B_{t_2} - B_{t_1}|$. \square

Now we let $\sigma \rightarrow 0$ and we consider the problem

$$\begin{cases} \partial_t m - \operatorname{div}(m Du B^{\varepsilon, N}(x)(B^{\varepsilon, N}(x))^T) = 0 & \text{in } \mathbb{R}^3 \times (0, T) \\ m(x, 0) = m_0(x) & \text{on } \mathbb{R}^3 \end{cases} \quad (4.11)$$

where u is the unique bounded solution to problem (4.2).

Lemma 4.4. *For N sufficiently large, problem (4.11) admits exactly one solution m_N in the space $C^{1/2}([0, T], \mathcal{P}_1(\mathbb{R}^3)) \cap L^\infty(0, T; \mathcal{P}_2(\mathbb{R}^3))$.*

Moreover, the solution m_N has a density in $L^\infty(\mathbb{R}^3 \times (0, T))$ and it is the image of the initial distribution through the flow

$$\dot{\gamma}(t) = -Du(\gamma(t), t)B^\varepsilon(\gamma(t))(B^\varepsilon(\gamma(t)))^T, \quad \gamma(0) = x \quad (4.12)$$

which is uniquely determined for m_0 -a.e. $x \in \mathbb{R}^3$.

Proof. Fix N sufficiently large (it will be suitably chosen later on). For any $\sigma \in (0, 1]$, let $(u_\sigma, m_{\sigma, N})$ be the unique classical bounded solution to system (4.4) (see Lemma 4.1 and Lemma 4.2). Letting $\sigma \rightarrow 0^+$, by the estimates in Lemma 4.1, in Lemma 4.2 and in Lemma 4.3, possibly passing to a subsequence that we still denote $(u_\sigma, m_{\sigma, N})$, we get that the functions u_σ converge locally uniformly to the unique solution u to (4.2) while $m_{\sigma, N}$ converge to a function $m_N \in C^{1/2}([0, T], \mathcal{P}_1(\mathbb{R}^3)) \cap L^\infty(0, T; \mathcal{P}_2(\mathbb{R}^3))$ in the $C^0([0, T], \mathcal{P}_1(\mathbb{R}^3))$ -topology and in the weak- $*$ topology of $L^\infty(\mathbb{R}^3 \times (0, T))$. By the same arguments as those in [16, Theorem 4.20 (proof)], in particular the uniform semiconcavity, we obtain that m_N is a solution to (4.11).

Let us now establish uniqueness and representation formula for the solution m_N . We observe that, by Lemma 4.1-(b), the drift verifies $\|Du B^{\varepsilon, N}(B^{\varepsilon, N})^T\|_\infty \leq CN^2$ and in particular condition [3, eq. (8.1.20)] is fulfilled. Hence, we can apply the superposition principle in [3, Theorem 8.2.1]: there exists a measure η_N on $\mathbb{R}^3 \times \Gamma$ such that

- 1) $m_N(t) = e_t \# \eta_N$ for all $t \in (0, T)$ (recall: $e_t(x, \gamma) = \gamma(t)$)
- 2) $\eta_N = \int_{\mathbb{R}^3} (\eta_N)_x dm_0(x)$ where, for m_0 -a.e. $x \in \mathbb{R}^3$, the measure $(\eta_N)_x$ is concentrated on the set of pairs $(x, \gamma) \in \mathbb{R}^3 \times \Gamma$ where γ solves

$$\dot{\gamma}(t) = -Du(\gamma(t), t)B^{\varepsilon, N}(\gamma(t))(B^{\varepsilon, N}(\gamma(t)))^T \quad \text{a.e. } t \in (0, T), \quad \gamma(0) = x. \quad (4.13)$$

We now claim that, for m_0 -a.e. $x \in \mathbb{R}^3$, the solutions to (4.13) coincide with those of (4.12).

The first two components of system (4.13) satisfy

$$\gamma_1' = -u_{x_1} + \psi_N(\gamma_2)u_{x_3}, \quad \gamma_2' = -u_{x_2} - \psi_N(\gamma_1)u_{x_3}.$$

Since the cut-off function ψ_N grows at most linearly, using arguments similar to those in the proof of Lemma 3.4-(a), we obtain that any solution γ to (4.13) is bounded uniformly in N . Since m_0 has compact support, there exists a positive constant k (independent of N) such that, for m_0 -a.e. $x \in \mathbb{R}^3$, any solution γ to (4.13) verifies $\gamma(t) \in [-k, k]^3$ for any $t \in (0, T)$. Hence, choosing $N \geq k$, we get that problem (4.13) coincides with (4.12) if $x \in \text{supp}(m_0)$.

It remains only to prove that, for m_0 -a.e. $x \in \mathbb{R}^3$, problem (4.12) admits exactly one solution. To this end, it is enough to invoke Lemma 3.4 and taking into account that $u(\cdot, 0)$ is Lipschitz continuous. \square

Now we let $N \rightarrow +\infty$ and we establish Theorem 4.1 exploiting that m_N have compact support independent on N which in turn is due to compactness of $\text{supp}(m_0)$.

Proof of Theorem 4.1. By Lemma 4.4 all the solutions m_N to (4.11) coincide if N is sufficiently large. Hence, passing to the limit as $N \rightarrow \infty$, we obtain that problem (4.1) admits a solution m in the space $C^{1/2}([0, T], \mathcal{P}_1(\mathbb{R}^3)) \cap L^\infty(0, T; \mathcal{P}_2(\mathbb{R}^3))$ with a density in $L^\infty(\mathbb{R}^3 \times (0, T))$. Finally, the estimates follow from the corresponding estimates in Lemma 4.3. \square

5. Proof of Proposition 2.1 and Theorem 2.1

Proof of Proposition 2.1. We achieve the proof through a fixed point argument as in [33, Theorem 1.1 (proof)] or in [2, Theorem 2.1] taking advantage of the results of Lemma 4.4. We sketch here some detail of the proof omitting the index ε since it is fixed. Consider the set

$$\mathcal{C} := \{m \in C^{1/2}([0, T], \mathcal{P}_1(\mathbb{R}^3)); m(0) = m_0\}$$

endowed with the norm of $C^0([0, T]; \mathcal{P}_1(\mathbb{R}^3))$. Observe that it is a nonempty closed and convex subset of $C^0([0, T]; \mathcal{P}_1(\mathbb{R}^3))$. We introduce a map \mathcal{T} as follows: to any $m \in \mathcal{C}$ we associate the solution u to problem (3.13) 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 (4.1).

By Theorem 4.1, 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 [16, 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 $C^0([0, T]; \mathcal{P}_1(\mathbb{R}^3))$ -topology as $n \rightarrow \infty$. The functions $\mathcal{T}(m_n)$ satisfy the estimates (4.3) of Theorem 4.1 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 [16, Lemma 5.7]), Theorem 4.1 ensures that the sequence $(\mathcal{T}(m_n))_n$ is uniformly bounded in $C^{1/2}([0, T]; \mathcal{P}_1)$ and $L^\infty(0, T; \mathcal{P}_2)$. Hence we obtain that, possibly for a

subsequence (still denoted by $\mathcal{T}(m_n)$), $\mathcal{T}(m_n)$ converges to some μ in the $C^0([0, T]; \mathcal{P}_1(\mathbb{R}^3))$ -topology.

Invoking Schauder fixed point Theorem, we accomplish the proof of the existence of a solution (u, m) of system (2.7).

To get (2.10) we note that, by Lemma 4.4 and Theorem 4.1, all the solution m_N of the truncated problem (4.11) coincide with m if N is sufficiently large. Hence still from Lemma 4.4, if (u, m) is a solution of (2.7), then for any function $\phi \in C_0^0(\mathbb{R}^3)$, we have

$$\int_{\mathbb{R}^3} \phi dm(t) = \int_{\mathbb{R}^3} \phi(\bar{\gamma}_x(t))m_0(x) dx \tag{5.1}$$

where $\bar{\gamma}_x$ is the solution of (2.11) and it is uniquely defined for any $x \in \mathbb{R}^3$.

Uniform estimates a) and b) come from Lemma 4.1 and Lemma 4.3 noting that $m_N = m$ for N sufficiently large. \square

Now, we can prove Theorem 2.1.

Proof of Theorem 2.1. 1. The uniform estimates for u_ε and for m_ε in Proposition 2.1 ensure that there exist two subsequences, which we will still denote u_ε and respectively m_ε such that, as $\varepsilon \rightarrow 0^+$, u_ε converge to some function u locally uniformly in (x, t) and m_ε converge to some $m \in C^0([0, T], \mathcal{P}_1(\mathbb{R}^3))$ in $C^0([0, T], \mathcal{P}_1(\mathbb{R}^3))$ -topology and in weak- $*$ - $L_{loc}^\infty(\mathbb{R}^3 \times (0, T))$ topology. In particular, we get $m(0) = m_0$ and we deduce that u is Lipschitz continuous in x , locally Lipschitz continuous in t , semiconcave in x and $Du_\varepsilon \rightarrow Du$ a.e. (because of the semiconcavity estimate in Proposition 2.1-(a) and [14, Theorem 3.3.3]).

Being a solution to (2.7)-(ii), the function m_ε fulfills

$$\int_0^T \int_{\mathbb{H}^1} m_\varepsilon(-\partial_t \varphi + D_{\mathcal{H}}u_\varepsilon \cdot D_{\mathcal{H}}\varphi) dx dt = 0 \quad \forall \varphi \in C_c^\infty(\mathbb{R}^3 \times (0, T)). \tag{5.2}$$

Passing to the limit as $\varepsilon \rightarrow 0^+$ we get that m is a solution to (1.1)-(ii). On the other hand, by standard stability results for viscosity solutions, we obtain that the function u is a viscosity solution to (2.7)-(i).

2. Consider the function m found in point (i). Since $t \rightarrow m_t$ is narrowly continuous, applying [3, Theorem 8.2.1], we get that there exists a probability measure η^* in $\mathbb{R}^3 \times \Gamma$ which satisfies points (i) and (ii) of [3, Theorem 8.2.1]. We denote $\eta \in \mathcal{P}(\Gamma)$ the measure on Γ defined as $\eta(A) := \eta^*(\mathbb{R}^3 \times A)$ for every $A \subset \Gamma$ measurable. We claim that η is a MFG equilibrium. Indeed, by [3, equation (8.2.1)], we have $e_0 \# \eta = m_0$ so $\eta \in \mathcal{P}_{m_0}(\Gamma)$. On the other hand, by [3, Theorem 8.2.1]-(i), η is supported on the curves solving (3.21). From Lemma 3.4 such curves are optimal, i.e. belong to the set $\Gamma^\eta[x]$, hence our claim is proved.

Let us now prove that (u, m) is a mild solution. By [3, Theorem 8.2.1], we have $m_t = e_t \# \eta$. Moreover, by Lemma 3.1, the function u found in point (i) is the value function associated to m as in Definition 2.3-(ii). In conclusion (u, m) is a mild solution to (1.1). \square

6. Generalizations

6.1. Some structures of Heisenberg type

In this section we generalize the previous results to some structures of Heisenberg type (see [11, Definition 18.1.1] for the precise definition and [11, Theorem 18.2.1] for a useful characterization). Throughout this section, the state space is \mathbb{R}^n , with $n \geq 1$, and, for any $x = (x_1, \dots, x_n)$, the matrix $B = B(x) \in M^{n \times m}$ (for some $m \leq n$) has the form

$$\begin{pmatrix} h_{11} & 0 & 0 & \dots & 0 \\ h_{21}(x_1) & h_{22}(x_1) & 0 & \dots & 0 \\ h_{31}(x_1, x_2) & h_{32}(x_1, x_2) & h_{33}(x_1, x_2) & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{m1}(x_1, \dots, x_{m-1}) & h_{m2}(x_1, \dots, x_{m-1}) & h_{m3}(x_1, \dots, x_{m-1}) & \dots & h_{mm}(x_1, \dots, x_{m-1}) \\ h_{(m+1)1}(x_1, \dots, x_m) & h_{(m+1)2}(x_1, \dots, x_m) & h_{(m+1)3}(x_1, \dots, x_m) & \dots & h_{(m+1)m}(x_1, \dots, x_m) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{n1}(x_1, \dots, x_m) & h_{n2}(x_1, \dots, x_m) & h_{n3}(x_1, \dots, x_m) & \dots & h_{nm}(x_1, \dots, x_m) \end{pmatrix}.$$

In other words, the coefficients of the matrix B fulfill:

$$\begin{cases} h_{11} \in \mathbb{R} \setminus \{0\}; & \text{for } i \in \{1, \dots, m\} \text{ and } j > i, \quad h_{ij} = 0; \\ \text{for } i \in \{2, \dots, m\}, & h_{ij}(x) = h_{ij}(x_1, \dots, x_{i-1}); \\ \text{for } i \in \{m + 1, \dots, n\}, & h_{ij}(x) = h_{ij}(x_1, \dots, x_m). \end{cases} \tag{6.1}$$

We require that the coefficients h_{ij} fulfill the following hypotheses

- (H4) $h_{ij} \in C^2(\mathbb{R}^n)$ are globally Lipschitz continuous (in particular they have an at most linear growth at infinity) with D^2h_{ij} bounded; the h_{ij} 's fulfill (6.1);
- (H5) there exists a constant $C > 0$ such that $\|h_{ij}\|_\infty \leq C$ for any $i = 1, \dots, \min\{n - 1, m\}$, $j \in \{1, \dots, i\}$;
- (H6) $\{x \in \mathbb{R}^n : h_{11}h_{22} \dots h_{mm} = 0\}$ has null measure in \mathbb{R}^n .

Let us observe that the boundedness assumption in (H5) only concerns the first $\min\{n - 1, m\}$ rows of B ; hence, when $n = m$, the coefficients of the m -th row can be unbounded, while, when $n > m$, the coefficients of any i -th row with $i > m$ can be unbounded. Let us also emphasize that we do not require that the columns of B fulfill the Hörmander condition and neither assumption 2) of [11, Theorem 18.2.1].

Now, the generic player in the MFG aims at choosing a control $\alpha = (\alpha_1, \dots, \alpha_m) \in L^2(0, T; \mathbb{R}^m)$ so to minimize the cost in (1.6) when its dynamics $x(\cdot) = (x_1(\cdot), x_2(\cdot), \dots, x_n(\cdot))$ obeys to the differential equation

$$x'_k(s) = \sum_{j=1}^{\min\{k,m\}} h_{kj}(x(s))\alpha_j(s) \quad \text{with } k = 1, \dots, n. \tag{6.2}$$

Corollary 6.1. Assume hypotheses (H1)-(H6). Then, the results of Theorem 2.1 hold true.

Proof. The proof is just an adaptation with some heavy calculations of the proof of Theorem 2.1 so we only describe the main changes.

As before, for $\varepsilon \in (0, 1]$, we introduce the approximating problem (2.7) with the matrix

$$B^\varepsilon := \begin{pmatrix} h_{11} & \dots & 0 & 0 & \dots & 0 \\ h_{21}(x_1) & \dots & 0 & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \dots & 0 \\ h_{m1}(x_1, \dots, x_{m-1}) & \dots & h_{mm}(x_1, \dots, x_{m-1}) & 0 & \dots & 0 \\ h_{(m+1)1}(x_1, \dots, x_m) & \dots & h_{(m+1)m}(x_1, \dots, x_m) & \varepsilon & \dots & 0 \\ \vdots & \ddots & \vdots & 0 & \ddots & 0 \\ h_{n1}(x_1, \dots, x_m) & \dots & h_{nm}(x_1, \dots, x_m) & 0 & \dots & \varepsilon \end{pmatrix} \in M^{n \times n}$$

where the first m columns contain the matrix B while the last $(n - m)$ columns have coefficients

$$h_{ij} = \varepsilon \quad \text{if } i = j \in \{m + 1, \dots, n\}, \quad h_{ij} = 0 \quad \text{otherwise.}$$

Explicitly, for $p = (p_1, \dots, p_n)$, the Hamiltonian H^ε and the drift $\partial_p H^\varepsilon$ are respectively

$$H^\varepsilon(x, p) = \frac{1}{2} \sum_{i=1}^m \left(\sum_{k=i}^n h_{ki} p_k \right)^2 + \frac{\varepsilon^2}{2} \sum_{i=m+1}^n p_i^2$$

$$\frac{\partial H^\varepsilon}{\partial p_j} = \begin{cases} \sum_{i=1}^j h_{ji} \left(\sum_{k=i}^n h_{ki} p_k \right) & \text{if } j \in \{1, \dots, m\} \\ \sum_{i=1}^m h_{ji} \left(\sum_{k=i}^n h_{ki} p_k \right) + \varepsilon^2 p_j & \text{if } j \in \{m + 1, \dots, n\}. \end{cases}$$

Assumption (H6) guarantees the uniqueness of the optimal trajectory after the initial time, following the same argument as in Proposition 3.3.

The rest of the proof relies on the structure of B , namely on these facts: the first m rows have a diagonal form, the coefficients of the i -th row, with $i \leq m$ only depend on x_1, \dots, x_{i-1} , the coefficients of the last $(n - m)$ rows only depend on x_1, \dots, x_m . Indeed, these properties permit to proceed iteratively on the first m coordinates and, afterwards, to study the last $(n - m)$ ones.

For the sake of completeness let us detail the adaptation of the optimal synthesis of Lemma 3.4. To prove point a) it suffices to establish that the solution to the differential equation (3.21) is bounded and Lipschitz continuous (note that also in this case Du is bounded). Indeed, the first m coordinates in (3.21) read

$$\gamma'_j = \sum_{i=1}^j h_{ji} \left(\sum_{k=i}^n h_{ki} u_{x_k} \right). \tag{6.3}$$

We split our arguments according to the fact that $n = m$ or $n > m$.

For $n = m$, the structure of B ensures that all the coefficients h_{ij} are independent of x_m . In particular, we deduce that the system of differential equations (6.3) with $j = 1, \dots, m - 1$ is independent of γ_m and, by (H5), gives an estimate of the form $|\xi'| \leq C(|\xi| + 1)$ for $\xi := (\gamma_1, \dots, \gamma_{m-1})$. By standard theory, we deduce that, for $i \in \{1, \dots, m - 1\}$, the curves γ_i , are bounded and Lipschitz continuous. Finally, we plug this bound in (6.3) with $j = m$ and we obtain that also γ_m is bounded and Lipschitz continuous.

For $n > m$, assumption (H5) ensures that h_{ij} are all bounded for $j \in \{1, \dots, m\}$. We deduce that in the system of (6.3) with $j = 1, \dots, m$ all the right hand sides are Lipschitz continuous in $(\gamma_1, \dots, \gamma_m)$. In particular we obtain that, for $j = 1, \dots, m$, all the γ_j are bounded and Lipschitz continuous. Afterwards, we consider the system of (6.3) with $j = m + 1, \dots, n$ and, again, we have that all the right hand sides are bounded. Hence, γ is bounded and Lipschitz continuous.

To get point b) of Lemma 3.4 we choose as increment $\xi = (\xi_1, \dots, \xi_n) \in \mathbb{R}^n$ such that $\xi_k = \sum_{j=1}^n h_{kj}^\varepsilon(x) \beta_j$ where the β_j 's are arbitrary and the $h_{kj}^\varepsilon(x)$'s are the coefficients of $B^\varepsilon(x)$, namely $B^\varepsilon = (h_{kj}^\varepsilon)_{kj}$. Now, the necessary conditions read:

$$\begin{cases} (i) & x' = p B^\varepsilon(x) B^\varepsilon(x)^T, \\ (ii) & p' = -\frac{D(|p B^\varepsilon(x)|^2)}{2} + Df(x, s) \\ (iii) & x(t) = x, \quad p(T) = -Dg(x(T)) \end{cases} \tag{6.4}$$

and

$$\alpha(s) = p(s) B^\varepsilon(x(s)), \quad \text{a.e on } [t, T]. \tag{6.5}$$

We observe that, by equations (6.2) and (6.5), there holds $\partial_{x_i} (|p B^\varepsilon(x(s))|^2) = 2p \partial_{x_i}(x'(s))$ and consequently, equation (6.4)-(ii) can be written as

$$\partial_{x_i} f(x(s), s) = p'_i(s) + p \partial_{x_i}(x'(s)) \quad \text{for } i = 1, \dots, n.$$

One can follow the same arguments as in the proof of point b) of Lemma 3.4: integrating by parts and using the last relation and the final condition in (6.4)-(iii), one gets that $Du_\varepsilon(x, t) B^\varepsilon(x) = -\alpha(t)$ which, together with assumption (H6), yields the uniqueness of the optimal trajectory. \square

Let us now give some applications of the above result.

Example 6.1. *d-dimensional Heisenberg group.* In the state space \mathbb{R}^{2d+1} , with $d \in \mathbb{N} \setminus \{0\}$, consider the matrix

$$B(x) = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ -x_{d+1} & \dots & -x_{2d} & x_1 & \dots & x_d \end{pmatrix}$$

where the first $2d$ rows contain the identity matrix. This matrix fulfills assumptions (H4)-(H6) and encompasses the matrix in (1.3) when $d = 1$.

Example 6.2. *Completely degenerate case.* In the state space \mathbb{R}^n , consider the matrix

$$B(x) = \begin{pmatrix} I_m \\ 0_{(n-m),m} \end{pmatrix}$$

where I_m is the identity matrix $m \times m$ while $0_{(n-m),m}$ is the null matrix $(n - m) \times m$. With this matrix, the generic player in the MFG controls only its first m coordinates. This matrix fulfills assumptions (H4)-(H6).

Example 6.3. *Grushin case.* In the state space \mathbb{R}^2 , consider the matrix

$$B(x) = \begin{pmatrix} 1 & 0 \\ 0 & x_1 \end{pmatrix}.$$

This matrix fulfills assumptions (H4)-(H6). In particular, Corollary 6.1 deals with MFG with Grushin dynamics with unbounded coefficients which was not encompassed among the cases coped in [33, Theorem 1.1].

Example 6.4. *Car parking* ([36], eq. 2.1). We consider the problem for modeling the position of a car through its steering angle $x_1(t)$ and the position of its centre $(x_2(t), x_3(t))$. The control problem is

$$\dot{x} = \begin{pmatrix} 1 & 0 \\ 0 & \cos x_1 \\ 0 & \sin x_1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix}.$$

Clearly, the system fulfills (H4)-(H6).

Remark 6.1. Note that our result holds also for more general structures that do not satisfy assumptions of [11, Theorem 18.2.1]; indeed, for instance, the case in [11, Example 18.1.4] satisfies our assumptions even if is not a H -type group.

Remark 6.2. Let us finally observe that the results in Corollary 6.1 can be further generalized with some slight modifications of the previous arguments. For instance, one can weaken assumption (H5) requiring only

(H5') for $j \in \{1, \dots, m\}$, h_{ij} has an at most linear growth with respect to (x_1, \dots, x_j) uniformly in x .

In this case the proof is similar to the one of Corollary 6.1. We only give some details on the proof that the any solution to (3.21) is bounded and Lipschitz continuous. Indeed equation (3.21) reads as the system formed by equations (6.3) with $j \in \{1, \dots, n\}$. We proceed iteratively on j . For $j = 1$, equation (6.3) has a right hand side which has an at most linear growth only in γ_1 uniformly in γ . We deduce that γ_1 is bounded and Lipschitz continuous. Afterwards, by these properties of γ_1 , we deduce that the right hand side of equation (6.3) with $j = 2$ has an at most linear growth only in γ_2 ; hence also γ_2 is bounded and Lipschitz continuous. Iteratively, we get that the whole γ is bounded and Lipschitz continuous.

We here give an example of a 4×3 matrix which satisfies (H5') but not (H5):

$$B(x) = \begin{pmatrix} 1 & 0 & 0 \\ h_{21}(x_1) & x_1 & 0 \\ x_1 + x_2 & x_1 + x_2 & x_1 + x_2 \\ x_1 + x_2 + x_3 & x_1 + x_2 + x_3 & x_1 + x_2 + x_3 \end{pmatrix}$$

where the function h_{21} is bounded and with at most a linear growth at infinity. Indeed assumption (H5) requires also that the terms h_{22} and h_{3i} , $i = 1, 2, 3$ are bounded.

6.2. Hamiltonian of the form $H(x, p) = |pB(x)|^\gamma$ with $\gamma \in [1, 2]$

Our results can be extended to any Hamiltonian of the form $H(x, p) = |pB(x)|^\gamma$ with $\gamma \in [1, 2]$. We can write explicitly the Hamiltonian

$$H(x, p) := ((p_1 - x_2 p_3)^2 + (p_2 + x_1 p_3)^2)^{\gamma/2}$$

and the drift term $\partial_p H(x, p)$ in the continuity equation is

$$\begin{aligned} \partial_p H(x, p) &= \gamma (|pB(x)|^2)^{\gamma/2-1} pB(x)B(x)^T \\ &= \gamma \frac{1}{(|pB(x)|^2)^{1-\gamma/2}} (p_1 - x_2 p_3, p_2 + x_1 p_3, -p_1 x_2 + p_2 x_1 + p_3(x_1^2 + x_2^2)). \end{aligned}$$

Since $\gamma \in [1, 2]$ this term has still a sublinear growth.

In this case in the cost functional for the associated control problem we have to replace the term $|\alpha|^2$ with $|\alpha|^\gamma$, where $\frac{1}{\gamma} + \frac{1}{\gamma'} = 1$.

The results on the associated optimal control, found in Section 3 for $\gamma = 2$, hold also in this case. In particular from the optimality conditions, the optimal trajectories $x^*(s)$ associated to the optimal control problem given in (3.10) for $\gamma = 2$ satisfy:

$$x' = \partial_p H(x, Du) = \gamma \frac{1}{|pB(x)|^{2-\gamma}} (u_{x_1} - x_2 u_{x_3}, u_{x_2} + x_1 u_{x_3}, -u_{x_1} x_2 + u_{x_2} x_1 + u_{x_3}(x_1^2 + x_2^2)). \tag{6.6}$$

A key point to obtain our main result is the optimal synthesis proved in Lemma 3.4 for the case $\gamma = 2$, in particular the Lipschitz continuity of the solution of system (3.21). In this case, since Du is still bounded, the first two components of the right hand side of (6.6) have sublinear growth. Hence we can argue as in the proof of point a) of the proof of Lemma 3.4 to obtain that there exists a constant C such that $\xi(s) := (x_1(s), x_2(s))$ satisfies $|\xi'| \leq C(|\xi|^\beta + 1)$ with $\beta < 1$. Then $\xi(s)$ is bounded and from the third component of (6.6) also $x_3(s)$ is bounded, then the curve $x(s)$ is Lipschitz continuous. Since the drift term in the continuity equation has a sublinear behaviour we can repeat the arguments of Section 4 to get also in this case the main result.

Acknowledgments

The first and the second authors are members of GNAMPA-INdAM and were partially supported also by the research project of the University of Padova ‘‘Mean-Field Games and

Nonlinear PDEs”, by the Fondazione Cariparo Project “Nonlinear Partial Differential Equations: Asymptotic Problems and Mean-Field Games” and by KAUST project OSR-2017-CRG6-3452.01. The third author has been partially funded by the ANR project ANR-16-CE40-0015-01.

References

- [1] Y. Achdou, I. Capuzzo Dolcetta, Approximation of solutions of Hamilton-Jacobi equations on the Heisenberg group, *Math. Model. Numer. Anal.* 42 (4) (2008) 565–591.
- [2] Y. Achdou, P. Mannucci, C. Marchi, N. Tchou, Deterministic mean field games with control on the acceleration, *NoDEA Nonlinear Differ. Equ. Appl.* 27 (3) (2020) 33.
- [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] Z. Balogh, A. Calogero, R. Pini, The Hopf-Lax formula in Carnot groups: a control theoretic approach, *Calc. Var. Partial Differ. Equ.* 49 (3–4) (2014) 1379–1414.
- [5] M. Bardi, I. Capuzzo Dolcetta, *Optimal Control and Viscosity Solutions of Hamilton-Jacobi Bellman Equations*, Systems and Control: Foundations and Applications, Birkhäuser, Boston, 1997.
- [6] A. Bellaïche, The tangent space in sub-Riemannian geometry, in: *Sub-Riemannian Geometry*, in: *Progr. Math.*, vol. 144, Birkhäuser, Basel, 1996, pp. 1–78.
- [7] 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.
- [8] A. Bensoussan, J. Frehse, P. Yam, *Mean Field Games and Mean Field Type Control Theory*, Springer Briefs in Mathematics, Springer, New York, 2013.
- [9] P. Besala, H. Ugowski, Some uniqueness theorems for solutions of parabolic and elliptic partial differential equations in unbounded regions, *Colloq. Math.* 20 (1969) 127–141.
- [10] 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.
- [11] A. Bonfiglioli, E. Lanconelli, F. Uguzzoni, *Stratified Lie Groups and Potential Theory for Their Sub-Laplacians*, Springer Monographs in Mathematics, Springer, Berlin, 2007.
- [12] P. Cannarsa, R. Capuani, Existence and uniqueness for mean field games with state constraints, in: *PDE Models for Multi-Agent Phenomena*, in: *Springer INdAM Ser.*, vol. 28, Springer, Cham, 2018, pp. 49–71.
- [13] P. Cannarsa, C. Mendico, Mild and weak solutions of mean field game problems for linear control systems, *Minimax Theory Appl.* 5 (2) (2020) 221–250.
- [14] 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.
- [15] L. Capogna, G. Citti, G. Rea, A subelliptic analogue of Aronson-Serrin’s Harnack inequality, *Math. Ann.* 357 (3) (2013) 1175–1198.
- [16] P. Cardaliaguet, Notes on Mean Field Games, from P.L. Lions lectures at College de France, 2012, available at <https://www.ceremade.dauphine.fr/~cardalia/MFG20130420.pdf>.
- [17] P. Cardaliaguet, Long time average of first order mean field games and weak KAM theory, *Dyn. Games Appl.* 3 (2013) 473–488.
- [18] 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.
- [19] F. Clarke, *Functional Analysis, Calculus of Variations and Optimal Control*, Graduate Text in Mathematics, vol. 264, Springer-Verlag, London, 2013.
- [20] J.M. Coron, *Control and Nonlinearity*, Mathematical Surveys and Monographs, vol. 136, American Mathematical Society, Providence, RI, 2007.
- [21] F. Da Lio, O. Ley, Uniqueness results for second-order Bellman-Isaacs equations under quadratic growth assumptions and application, *SIAM J. Control Optim.* 45 (1) (2006) 74–106.
- [22] M. Defoort, T. Floquet, W. Perruquetti, S.V. Drakunov, Integral sliding mode control of an extended Heisenberg system, *IET Control Theory Appl.* 3 (10) (2009) 1409–1424.
- [23] F. Dragoni, E. Feleqi, Ergodic Mean Field Games with Hörmander diffusions, *Calc. Var. Partial Differ. Equ.* 57 (5) (2018) 116.
- [24] E. Feleqi, D. Gomes, T. Tada, Hypoelliptic mean field games- a case study, *Minimax Theory Appl.* 5 (2) (2020) 305–326.
- [25] D. Gomes, E.A. Pimentel, V. Voskanyan, *Regularity Theory for Mean-Field Game Systems*, SpringerBriefs in Mathematics, Springer, Berlin, 2016.

- [26] 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.
- [27] I. Karatzas, S.E. Shreve, *Brownian Motion and Stochastic Calculus*, second edition, Graduate Texts in Mathematics, vol. 113, Springer-Verlag, New York, 1991.
- [28] O.A. Ladyženskaja, V.A. Solonnikov, N.N. Ural'ceva, *Linear and Quasilinear Equations of Parabolic Type*, translated from the Russian by S. Smith, Translations of Mathematical Monographs, vol. 23, American Mathematical Society, Providence, RI, 1968.
- [29] J.-M. Lasry, P.-L. Lions, Jeux à champ moyen. I. Le cas stationnaire, *C. R. Math. Acad. Sci. Paris* 343 (2006) 619–625.
- [30] 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.
- [31] J.-M. Lasry, P.-L. Lions, Mean field games, *Jpn. J. Math. New Ser.* 2 (2007) 229–260.
- [32] G.M. Lieberman, *Second Order Parabolic Differential Equations*, World Scientific Publishing Co., Inc., River Edge, NJ, 1996.
- [33] P. Mannucci, C. Mariconda, C. Marchi, N. Tchou, Non-coercive first order Mean Field Games, *J. Differ. Equ.* 269 (5) (2020) 4503–4543.
- [34] R. Montgomery, *A Tour of SubRiemannian Geometries, Their Geodesics and Applications*, AMS, Providence, RI, 2002.
- [35] R.M. Murray, S.S. Sastry, Nonholonomic motion planning: steering using sinusoids, *IEEE Trans. Autom. Control* 38 (5) (1993) 700–716.
- [36] J.A. Reeds, L.A. Shepp, Optimal paths for a car that goes both forwards and backwards, *Pac. J. Math.* 145 (2) (1990) 367–393.