SBV REGULARITY FOR HAMILTON-JACOBI EQUATIONS WITH HAMILTONIAN DEPENDING ON (T, X).

STEFANO BIANCHINI * AND DANIELA TONON †

Abstract. In this paper we prove the SBV regularity of the gradient of a viscosity solution of the Hamilton-Jacobi equation

$$\partial_t u + H(t, x, D_x u) = 0 \quad \text{in } \Omega \subset [0, T] \times \mathbb{R}^n,$$

under the hypothesis of uniform convexity of the Hamiltonian H in the last variable. This result extends the result of Bianchini, De Lellis and Robyr obtained for a Hamiltonian $H = H(D_x u)$ which depends only on the spatial gradient of the solution.

Key words. SBV regularity, viscosity solutions, Hamilton-Jacobi equations

AMS subject classifications. 35F21, 49L25, 49N60

1. Introduction. We consider a viscosity solution u to the Hamilton-Jacobi equation

$$\partial_t u + H(t, x, D_x u) = 0 \qquad \text{in } \Omega \subset [0, T] \times \mathbb{R}^n. \tag{1.1}$$

It is well known that even when the initial datum for (1.1) is extremely regular, the viscosity solution of the Cauchy problem develops singularities of the gradient in finite time. The structure of the non-differentiability set of the viscosity solution has been studied by several authors, see for example Fleming [11], Cannarsa and Soner [8]. As a major assumption they restrict to the case where the Hamiltonian $H(t, x, p)$ is strictly convex with respect to p and smooth in all variables. Under this restriction the viscosity solution of (1.1) can be represented as the value function of a classical problem in calculus of variation and it is semiconcave, see [7]. The semiconcavity of u ensures that u is twice differentiable almost everywhere and that its distributional Hessian is a measure with locally bounded variation. However, deeper results on its regularity have been proved. A significant result in our direction was obtained by Cannarsa, Mennucci and Sinestrari in [6]: they proved the SBV regularity of the gradient of the viscosity solution u, when u is the solution of the Cauchy problem with a regular initial datum $u(0, \cdot) = u_0(\cdot)$ belonging to $W^{1,\infty}(\mathbb{R}^n) \cap C^{R+1}(\mathbb{R}^n)$, with $R \geq 1$. Furthermore, they gave a sharper estimate on the set of regular conjugate points, which implies in particular that this set has Hausdorff dimension less than $n - 1$ if the initial datum is C^∞. Thus in particular they proved that the closure of the set of irregular points is $\mathcal{H}^n$-rectifiable.

Motivated by the work of Bianchini, De Lellis and Robyr in [5], we prove the SBV regularity for the gradient of the viscosity solution, reducing the regularity of the initial datum. Indeed, in that paper, the authors proved that the gradient of a viscosity solution of

$$\partial_t u + H(D_x u) = 0 \quad \text{in } \Omega \subset [0, T] \times \mathbb{R}^n$$

*SISSA, via Bonomea, 265, 34136 Trieste, Italy, Phone +39 040 3787 434, Fax +39 040 3787 528(`bianchin@sissa.it`).

†Institut de mathématiques de Jussieu, Université Pierre et Marie Curie, Paris 6, 4 Place de Jussieu, 75252 Paris Cedex 05, Phone +33 144279863 (`tonondaniela83@gmail.com`).

belongs locally to SBV under the assumption of uniform convexity of the Hamiltonian. This last assumption is stronger than the one of strict convexity used in [6], however the regularity of the initial datum is required to be only bounded and Lipschitz.

To be more precise we prove the SBV regularity of $D_x u$ and $\partial_t u$ under the following hypotheses

(H1) $H \in C^3([0,T] \times \mathbb{R}^n \times \mathbb{R}^n)$ with bounded second derivatives and there exist positive constants a, b, c such that for every $t \in [0,T], x \in \mathbb{R}^n, p \in \mathbb{R}^n$
  i) $H(t,x,p) \geq -c$,
  ii) $H(t,x,0) \leq c$,
  iii) $|H_{px}(t,x,p)| \leq a + b|p|$,

(H2) there exists $c_H > 0$ such that

$$c_H^{-1} Id_n(p) \leq H_{pp}(t,x,p) \leq c_H Id_n(p)$$

for any t, x.

The main theorem of the paper is the following.

Theorem 1.1. *Let u be a viscosity solution of (1.1), assume* (H1), (H2) *and set* $\Omega_t := \{x \in \mathbb{R}^n | \ (t,x) \in \Omega\}$. *Then the set of times*

$$S := \{t \mid D_x u(t,\cdot) \notin [SBV_{loc}(\Omega_t)]^n\}$$

is at most countable. In particular $D_x u \in [SBV_{loc}(\Omega)]^n, \partial_t u \in SBV_{loc}(\Omega)$.

Moreover, under the hypotheses

(H1-bis) $H \in C^3(\mathbb{R}^n \times \mathbb{R}^n)$ with bounded second derivatives and there exist positive constants a, b, c such that for every $x \in \mathbb{R}^n, p \in \mathbb{R}^n$
  i) $H(x,p) \geq -c$,
  ii) $H(x,0) \leq c$,
  iii) $|H_{px}(x,p)| \leq a + b|p|$,

(H2-bis) there exists $c_H > 0$ such that

$$c_H^{-1} Id_n(p) \leq H_{pp}(x,p) \leq c_H Id_n(p)$$

for any x,

as a consequence of the theorem above, we have the following corollary.

Corollary 1.2. *Under assumptions (H1-bis), (H2-bis), the gradient of any viscosity solution u of*

$$H(x, Du) = 0 \qquad \text{in } \Omega \subset \mathbb{R}^n,$$

belongs to $[SBV_{loc}(\Omega)]^n$.

When the Hamiltonian of the previous corollary depends also on u, we can expect the SBV regularity of Du as long as the viscosity solution u is semiconcave. This happens for example when the Hamiltonian $H = H(x,u,p)$ is semiconcave with respect to x with constant independent from u and p, see [7].

The motivation for studying the SBV regularity of a function arises from problems in Control Theory and from image segmentation and measure-theoretic questions. The fact that the solution u does not have Cantor part in the second derivative implies that its structure is much simpler: it can be thought as almost piecewise $W^{2,1}$, with jumps in the gradient Du only along Lipschitz curves. In general, when studying perturbations of the solution, the Cantor part of the gradient is an instable structure, while the absolutely continuous part and the jump part do not disappear. Roughly speaking, this instability is the reason why the nonlinearity of H makes the Cantor part to disappear.

In Section 2 we recall preliminary results and definitions necessary to understand the main theorem. In Section 3 we show the properties of the unique viscosity solution to our Hamilton-Jacobi equation, we define generalized backward characteristics and we prove their no-crossing property. Finally in Section 4 we prove all the necessary lemmas and the main theorem.

2. Preliminaries.

2.1. Generalized differentials. We begin with the definition of generalized differential, see Cannarsa and Sinestrari [7] and Cannarsa and Soner [8].

Let Ω be an open subset of $\mathbb{R}^n$.

DEFINITION 2.1. *Let $u : \Omega \to \mathbb{R}$, for any $x \in \Omega$ the sets*

$$D^-u(x) = \left\{ p \in \mathbb{R}^n | \ \liminf_{y\to x} \frac{u(y) - u(x) - \langle p, y - x\rangle}{|y - x|} \geq 0 \right\},$$

$$D^+u(x) = \left\{ p \in \mathbb{R}^n | \ \limsup_{y\to x} \frac{u(y) - u(x) - \langle p, y - x\rangle}{|y - x|} \leq 0 \right\},$$

are called, respectively, the subdifferential *and* superdifferential *of u at x.*

DEFINITION 2.2. *Let $u : \Omega \to \mathbb{R}$ be locally Lipschitz. A vector $p \in \mathbb{R}^n$ is called a* reachable gradient *of u at $x \in \Omega$ if there exists a sequence $\{x_k\} \subset \Omega \setminus \{x\}$ such that u is differentiable at x_k for each $k \in \mathbb{N}$, and*

$$\lim_{k\to\infty} x_k = x, \qquad \lim_{k\to\infty} Du(x_k) = p.$$

The set of all reachable gradients of u at x is denoted by $D^*u(x)$.

2.2. BV and SBV functions. A detailed description of the spaces BV and SBV can be found in Ambrosio, Fusco and Pallara [3], Chapters 3 and 4. For the reader convenience, we briefly recall that, given $u \in BV(\mathbb{R}^n)$, it is possible to decompose the distributional derivative of u, which by definition must be a measure with bounded total variation, into three mutually singular measures:

$$Du = D_a u + D_c u + D_j u.$$

$D_a u$ is the absolutely continuous part with respect to the Lebesgue measure. $D_j u$ is the part of the measure which is concentrated on the rectifiable $n-1$ dimensional set J, where the function u has jump discontinuities, thus for this reason it is called jump part. $D_c u$, the Cantor part, is the singular part which satisfies $D_c u(E) = 0$ for every Borel set E with $\mathcal{H}^{n-1}(E) < \infty$. If this part vanishes, i.e. $D_c u = 0$, we say that $u \in SBV(\mathbb{R}^n)$. When $u \in [BV(\mathbb{R}^n)]^k$ the distributional derivative Du is a matrix of Radon measures and the decomposition can be applied to every component of the matrix.

2.3. Semiconcave functions. For a complete introduction to the theory of semiconcave functions we refer to Cannarsa and Sinestrari [7], Chapter 2 and 3 and Lions [14]. For our purpose we define semiconcave functions with a linear modulus of semiconcavity. In general this class is considered only as a particular subspace of the class of semiconcave functions with general semiconcavity modulus. The proofs of the following statements can be found in the mentioned references.

Definition 2.3. *We say that a function $u : \Omega \to \mathbb{R}$ is* semiconcave *and we denote with $SC(\Omega)$ the space of functions with such a property, if $\exists C > 0$ such that for any $x, z \in \Omega$ such that the segment $[x - z, x + z]$ is contained in Ω*

$$u(x + z) + u(x - z) - 2u(x) \leq C|z|^2.$$

Proposition 2.4. *Let $u : \Omega \to \mathbb{R}$ belongs to $SC(\Omega)$ with semiconcavity constant $C \geq 0$. Then the function*

$$\tilde{u} : x \mapsto u(x) - \frac{C}{2}|x|^2$$

is concave, i.e. for any x, y in Ω such that the whole segment $[x, y]$ is contained in Ω, $\lambda \in [0, 1]$

$$\tilde{u}(\lambda x + (1 - \lambda)y) \geq \lambda\tilde{u}(x) + (1 - \lambda)\tilde{u}(y).$$

Theorem 2.5. *Let $u : \Omega \to \mathbb{R}$ belongs to $SC(\Omega)$. Then the following properties hold.*

i) (Alexandroff's Theorem) u is twice differentiable $\mathcal{H}^n$-a.e.; that is, for $\mathcal{H}^n$-a.e. $x_0 \in \Omega$, there exist a vector $p \in \mathbb{R}^n$ and a symmetric matrix M such that

$$\lim_{x \to x_0} \frac{u(x) - u(x_0) - \langle p, x - x_0\rangle + \langle M(x - x_0), x - x_0\rangle}{|x - x_0|^2} = 0.$$

ii) The gradient of u, defined a.e. in Ω, belongs to the class $BV_{loc}(\Omega, \mathbb{R}^n)$.

iii) Let $x \in \Omega$ then

$$D^+u(x) = \mathrm{co}D^*u(x),$$

where $\mathrm{co}A := \min\{B \subset \mathbb{R}^n \mid B \supset A, B \text{ convex}\}$ is the convex hull of $A \subset \mathbb{R}^n$. Thus D^+u is non empty at each point. Moreover D^+u is upper semicontinuous. See [8].

iv) The function $T(x) := -D^+\tilde{u}(x)$ is a maximal monotone function, i.e.

$$\langle y_1 - y_2, x_1 - x_2\rangle \geq 0 \qquad \forall x_i \in \Omega,\ y_i \in T(x_i),\ i = 1, 2;$$

and it is maximal in following sense

$$V \supset T,\ V \text{ monotone} \Longrightarrow V = T.$$

As stated in the above theorem at point ii), when u is semiconcave Du is a BV map, hence the distributional Hessian D^2u is a symmetric matrix of Radon measures and can be split into the three mutually singular parts $D^2_a u, D^2_j u, D^2_c u$. Moreover the following proposition holds.

Proposition 2.6. *Let u be a semiconcave function. If D denotes the set of points where D^+u is not single-valued, then $|D^2_c u|(D) = 0$.*

Proof. Indeed, the set of points where D^+u is not single-valued, i.e. the set of singular points, is a $\mathcal{H}^{n-1}$-rectifiable set. □

Definition 2.7. *We say that a function $v : \Omega \to \mathbb{R}$ is* semiconvex *if $u := -v$ is semiconcave.*

2.4. Viscosity solutions. A concept of generalized solutions to the equations

$$\partial_t u + H(t, x, D_x u) = 0 \qquad \text{in } \Omega \subset [0, T] \times \mathbb{R}^n, \tag{2.1}$$

and

$$H(x, Du) = 0 \qquad \text{in } \Omega \subset \mathbb{R}^n, \tag{2.2}$$

was found to be necessary since classical solutions break down and solutions which satisfy (2.1) almost everywhere are not unique. Crandall and Lions introduced in [10] the notion of viscosity solution to solve both these problems, see also Crandall, Evans and Lions [9].

DEFINITION 2.8. *A bounded uniformly continuous function $u : \Omega \to \mathbb{R}$ is called a* viscosity solution *of (2.1) (resp. (2.2)) provided that*

i) u is a viscosity subsolution of (2.1) (resp. (2.2)): for each $v \in C^\infty(\Omega)$ such that $u - v$ has a maximum at $(t_0, x_0) \in \Omega$ (resp. $x_0 \in \Omega$),

$$\partial_t v(t_0, x_0) + H(t_0, x_0, D_x v(t_0, x_0)) \leq 0 \qquad (\text{resp. } H(x_0, Dv(x_0)) \leq 0);$$

ii) u is a viscosity supersolution of (2.1) (resp. (2.2)): for each $v \in C^\infty(\Omega)$ such that $u - v$ has a minimum at $(t_0, x_0) \in \Omega$ (resp. $x_0 \in \Omega$),

$$\partial_t v(t_0, x_0) + H(t_0, x_0, D_x v(t_0, x_0)) \geq 0 \qquad (\text{resp. } H(x_0, Dv(x_0)) \geq 0).$$

3. Properties of the viscosity solution of Hamilton-Jacobi equations. The proofs of the following statements can be found in Cannarsa and Sinestrari [7], Chapter 6. See also Fleming [11], Fleming and Rishel [12], Fleming and Soner [13] and Lions [14].

We will consider here only viscosity solutions of equation (2.1), similar results apply also to viscosity solutions of the Hamilton-Jacobi equation (2.2).

The convexity of the Hamiltonian in the p-variable relates Hamilton-Jacobi equations to a variational problem.

Let L be the Lagrangian of our system, i.e. the Legendre transform of the Hamiltonian H with respect to the last variable, for any t, x fixed

$$L(t, x, v) := \sup_p \{\langle v, p \rangle - H(t, x, p)\}.$$

The Legendre transform inherits the properties of H, in particular L is $C^3([0, T] \times \mathbb{R}^n \times \mathbb{R}^n)$ and uniformly convex in the last variable.

In addition to the uniform convexity and C^3 regularity of L, the hypotheses on H, (H1) and (H2), ensure the existence of positive constants a, b, c such that for any $t \in [0, T], x \in \mathbb{R}^n, v \in \mathbb{R}^n$

i) $L(t, x, v) \geq -c$,
ii) $L_x(t, x, 0) \leq c$,
iii) $|L_{vx}(t, x, v)| \leq a + b|v|$.

Define the value function $u(\cdot, \cdot)$ associated the the bounded Lipschitz function $u_0(\cdot)$, for $(t, x) \in \Omega$

$$u(t, x) := \min \left\{ u_0(\xi(0)) + \int_0^t L(s, \xi(s), \dot{\xi}(s)) ds \;\middle|\; \xi(t) = x,\ \xi \in [C^2([0, t])]^n \right\}. \tag{3.1}$$

Less regularity can be asked to ξ, but it is unnecessary since any minimizing curve exists and is smooth, due to the regularity of L, see [7].

THEOREM 3.1. *Taken a minimizing curve ξ in (3.1), for the point (t, x), such that $\xi(s) \in \Omega_s$ for all $s \in [0, t]$, the following holds. (Recall $\Omega_s = \{x \in \mathbb{R}^n |\ (s, x) \in \Omega\}$.)*

i) The map $s \mapsto L_v(s,\xi(s),\dot\xi(s))$ is absolutely continuous.

ii) ξ is a classical solution to the Euler-Lagrange equation

$$\frac{d}{ds}L_v(s,\xi(s),\dot\xi(s)) = L_x(s,\xi(s),\dot\xi(s)),$$

and to the Du Bois-Reymond equation

$$\frac{d}{ds}[L(s,\xi(s),\dot\xi(s)) - \langle \dot\xi(s), L_v(s,\xi(s),\dot\xi(s))\rangle] = L_t(s,\xi(s),\dot\xi(s)),$$

for all $s \in [0,t]$, where $L_t(s,\xi(s),\dot\xi(s))$ is the derivative of L with respect to the first variable.

iii) For any $r > 0$ there exists $K(r) > 0$ such that, if $(t,x) \in [0,r] \times B_r(0)$, then

$$\sup_{s\in[0,t]} |\dot\xi(s)| \le K(r).$$

iv) There exists a dual arc or co-state

$$p(s) := L_v(s,\xi(s),\dot\xi(s)) \qquad s \in [0,t], \tag{3.2}$$

such that ξ, p solve the following system

$$\begin{cases} \dot\xi(s) = H_p(s,\xi(s),p(s)) \\ \dot p(s) = -H_x(s,\xi(s),p(s)). \end{cases}$$

v) $(s,\xi(s))$ is regular, i.e. for any $0 < s < t$ ξ is the unique minimizer for $u(s,\xi(s))$, and $u(s,\cdot)$ is differentiable at $\xi(s)$.

vi) Let p be the dual arc associated to ξ as in (3.2) then we have

$$p(t) \in D^+_x u(t,x),$$

$$p(s) = D_x u(s,\xi(s)), \qquad s \in (0,t).$$

Theorem 3.2. *The value function u defined in (3.1) is a viscosity solution of (2.1) with bounded Lipschitz initial datum*

$$u(0,x) = u_0(x).$$

We present below some properties of the unique viscosity solution to the Hamilton-Jacobi equation (2.1), which follow from the representation formula we have just seen. These properties are taken from [7].

Theorem 3.3 (Dynamic Programming Principle). *Fix $(t,x) \in \Omega$, then for all $t' \in [0,t)$*

$$u(t,x) := \min\left\{ u(t',\xi(t')) + \int_{t'}^{t} L(s,\xi(s),\dot\xi(s))ds \;\middle|\; \xi(t) = x,\ \xi \in [C^2([t',t])]^n \right\}. \tag{3.3}$$

Moreover, if ξ is a minimizer in (3.1) it is a minimizer also in (3.3), for any $t' \in [0,t)$.

Theorem 3.4 (Semiconcavity Theorem). *Suppose* (H1), (H2) *hold and u_0 belongs to $C_b(\mathbb{R}^n)$. Then for any t in $(0,T]$, $u(t,\cdot)$ is locally semiconcave with semiconcavity constant $C(t) = \frac{C}{t}$. Thus for any fixed $\tau > 0$ there exists a constant $C = C(\tau)$ such that $u(t,\cdot)$ is semiconcave with constant less than C for any $t \ge \tau$.*

Moreover, u is also locally semiconcave in both the variables (t,x) in Ω.

3.1. Minimizers and Generalized Backward Characteristics. We introduce the definition of generalized backward characteristics.

DEFINITION 3.5. *Given $x \in \Omega_t$ for t fixed in $[0, T]$, we call* generalized backward characteristic*, associated to u starting from x, the curve $s \mapsto (s, \xi(s))$, where $\xi(\cdot)$ and its dual arc $p(\cdot)$ solve the system*

$$\begin{cases} \dot{\xi}(s) = H_p(s, \xi(s), p(s)) \\ \dot{p}(s) = -H_x(s, \xi(s), p(s)) \end{cases} \tag{3.4}$$

with final conditions

$$\begin{cases} \xi(t) = x \\ p(t) = p, \end{cases} \tag{3.5}$$

where $p \in D^+_x u(t, x)$.

If $D^+_x u(t, x)$ is single-valued then we call ξ a classical backward characteristic.

We state here some properties of minimizers which strictly relate them with classical and generalized characteristics, see [7].

THEOREM 3.6. *For any $(t, x) \in \Omega$ the map that associates with any $(p_t, p_x) \in D^*u(t, x)$ the curve ξ obtained by solving the system (3.4) with the final conditions*

$$\begin{cases} \xi(t) = x \\ p(t) = p_x \end{cases}$$

provides a one-to-one correspondence between $D^*u(t, x)$ and the set of minimizers of $u(t, x)$.

Thus we can state the following theorem which follows from Theorem 3.1-(iv), Theorem 3.6 and Definition 3.5.

THEOREM 3.7. *Let (t, x) in Ω be given, and let ξ be a C^2 curve such that $\xi(s) \in \Omega_s$ for all $0 \leq s \leq t$.*

Then ξ is a minimizer if and only if ξ and its dual arc p are solutions of the system (3.4), for any $s \in [0, t]$, with the final conditions (3.5), where $(-H(t, x, p), p)$ belongs to $D^*u(t, x)$.

A minimizer ξ is a generalized backward characteristic. In particular ξ is a classical backward characteristic if and only if ξ is the unique minimizer for $u(t, x)$. The set of minimizers for $u(t, x)$ can be a proper *subset of the set of generalized backward characteristics emanated from (t, x).*

REMARK 3.8. *Note that, the solutions ξ of the system (3.4) are in general curves and not straight lines, as solutions were in the case $H = H(p)$.*

REMARK 3.9. No-crossing property of minimizers. *Fix a time t and consider a minimizing curve ξ such that $\xi(t) = x \in \Omega_t$. For $0 < s < t$ the curve ξ is the unique minimizer for $u(s, \xi(s))$, this ensures that any other minimizer cannot intersect ξ for any $0 < s < t$ (otherwise uniqueness would be lost, see point* (v) *of Theorem 3.1). As a consequence generalized backward characteristics which are also minimizers, i.e. solution of (3.4), (3.5), where $(-H(t, x, p), p)$ belongs to $D^*u(t, x)$, cannot intersect except at time 0 or t. Nothing can be said at this level for generalized backward characteristics solution to (3.4) with*

$$\xi(t) = x \qquad p(t) = p \in D^+_x u(t, x) \setminus D^*_x u(t, x),$$

which are not minimizers. In general they cross and they are not regular.

The introduction of a *backward solution*, as in Barron, Cannarsa, Jensen and Sinestrari [4], will allow us to see that, at least for a small interval of time, all the generalized backward characteristics share the no-crossing property.

Fix t in $(0,T]$ and define for $0 \le \tau < t$, $y \in \Omega_\tau$ the function

$$\tilde{u}(\tau,y) := \max\left\{ u(t,\xi(t)) - \int_\tau^t L(s,\xi(s),\dot{\xi}(s))ds \;\middle|\; \xi(\tau)=y,\ \xi\in[C^2([\tau,t])]^n \right\}. \tag{3.6}$$

Note that, as explained in [4], the function $v(\tau,y) := \tilde{u}(t-\tau,y)$ is a viscosity solution of

$$\partial_\tau v - H(t-\tau,y,D_y v) = 0 \qquad \text{in } \Omega\subset[0,T]\times\mathbb{R}^n$$

with initial datum $v(0,y) = \tilde{u}(t,y) = u(t,y)$, for this reason $\tilde{u}$ is called backward solution.

Proposition 3.10. *In general*

$$\tilde{u}(\tau,y) \le u(\tau,y)$$

and the equality holds if and only if the maximizer ξ in (3.6), defined for $\tau\le s\le t$, is part of a minimizing curve for $u(t,\xi(t))$.

Proof. Let ξ be a C^2-curve which is a maximizer for $\tilde{u}(\tau,y)$, i.e.

$$\tilde{u}(\tau,y) = u(t,\xi(t)) - \int_\tau^t L(s,\xi(s),\dot{\xi}(s))ds.$$

Thanks to the Dynamic Programming Principle,

$$u(t,\xi(t)) \le u(\tau,y) + \int_\tau^t L(s,\xi(s),\dot{\xi}(s))ds.$$

Hence,

$$\tilde{u}(\tau,y) \le u(\tau,y)$$

and the equality holds if and only if ξ is also a minimizer for $u(t,\xi(t))$, thus $D_x^+u(s,\xi(s))$ is single-valued for any $\tau\le s<t$. □

Note that a curve ξ which is a minimizer for $u(t,x)$ is also a maximizer for $\tilde{u}(\tau,\xi(\tau)) = u(\tau,\xi(\tau))$ for any $0\le\tau<t$.

With suitable modifications Theorems 3.1, 3.2, 3.3 and 3.4 still hold for $\tilde{u}(\tau,y)$ and its maximizers, in particular $\tilde{u}$ is semiconvex (rather than semiconcave) with constant $\frac{C}{t-\tau}$.

Without adding any other assumption, the no-crossing property holds also for maximizers.

However, if we restrict to a τ which is not too far from t, we can establish a one to one correspondence between generalized backward characteristics, as in Definition 3.5, and maximizers of (3.6), thus obtaining regularity and the no-crossing property for generalized backward characteristics. Moreover the backward solution $\tilde{u}(s,\cdot)$ belongs to $C^{1,1}(\Omega_s)$ for every $s\in(\tau,t)$.

To prove the above fact let us first reduce to a simpler case which will be useful also later on, during the proof of our main theorem.

LEMMA 3.11. *Consider the solutions to the system (3.4) with final conditions*

$$\begin{cases} \xi(t) = x \\ p(t) = p \in K \end{cases} \tag{3.7}$$

where x is fixed in $\mathbb{R}^n$ and K is a compact set in $\mathbb{R}^n$. For $t-\tau$ small enough there exists a one to one correspondence between p in K and $\xi(\tau)$ when $\xi(\cdot)$ is a solution of (3.4),(3.7).

Proof. Thanks to the Taylor expansion of the flow generated by (3.4), the solution to that system, with (3.7) as final conditions, is equal to

$$\xi(\tau) = x - (t-\tau)H_p(t,x,p) + O((t-\tau)^2),$$

and differentiating in p

$$\xi_p(\tau) = -(t-\tau)H_{pp}(t,x,p) + O((t-\tau)^2). \tag{3.8}$$

Note that ξ_p and p_p satisfy

$$\begin{cases} \dot{\xi}_p(s) = H_{px}(s,\xi(s),p(s))\xi_p(s) + H_{pp}(s,\xi(s),p(s))p_p(s) \\ \dot{p}_p(s) = -H_{xx}(s,\xi(s),p(s))\xi_p(s) - H_{xp}(s,\xi(s),p(s))p_p(s) \end{cases}$$

with final conditions

$$\begin{cases} \xi_p(t) = 0 \\ p_p(t) = Id_n(p). \end{cases}$$

Since the coefficients are smooth, equation (3.8) is precisely the Taylor expansion of $\xi_p(\tau)$.

Call $\omega := \frac{x-\xi(\tau)}{t-\tau}$. We have that ω_p is uniformly different from zero since

$$\omega_p = H_{pp}(t,x,p) + O(t-\tau).$$

Thus, restricting to $t-\tau$ small enough, we can locally invert this equation and obtain

$$p_\omega = L_{vv}\,(t,x,\omega) + O(t-\tau). \tag{3.9}$$

Moreover, from

$$\omega = H_p(t,x,p) + O(t-\tau),$$

integrating (3.9), we obtain

$$p = L_v\,(t,x,\omega) + O(t-\tau).$$

Thus we have reached a one to one correspondence between $\xi(\tau)$ and the value p of its dual curve at time t. □

Integrating (3.8) in p between p_1 and p_2 we obtain

$$\frac{\xi_1(\tau)-\xi_2(\tau)}{\tau - t} = H_p(t,x,p_1) - H_p(t,x,p_2) + O(t-\tau)(p_1-p_2)$$

where ξ_1 and ξ_2 are the generalized backward characteristics with initial data p_1 and p_2 respectively.

Proposition 3.12. *Consider a solution ξ to the system (3.4) with final conditions (3.7), let $y := \xi(\tau)$ and consider the straight line joining x to y*

$$\eta(s) = \frac{s-\tau}{t-\tau}x + \frac{t-s}{t-\tau}y. \tag{3.10}$$

Then we have the following estimates

$$\|\eta - \xi\|_{[C^0([\tau,t])]^n}, \|\eta_p - \xi_p\|_{[C^0([\tau,t])]^{n^2}}, \|\eta_{pp} - \xi_{pp}\|_{[C^0([\tau,t])]^{n^3}} \leq O((t-\tau)^2),$$

$$\|\dot\eta - \dot\xi\|_{[C^0([\tau,t])]^n}, \|\dot\eta_p - \dot\xi_p\|_{[C^0([\tau,t])]^{n^2}}, \|\dot\eta_{pp} - \dot\xi_{pp}\|_{[C^0([\tau,t])]^{n^3}} \leq O(t-\tau).$$

Proof. As we saw in the previous proposition

$$y = \xi(\tau) = x - (t-\tau)H_p(t,x,p) + O((t-\tau)^2),$$

and for $s \in [\tau, t]$

$$\xi(s) = x - (t-s)H_p(t,x,p) + O((t-s)^2).$$

Compute now the difference

$$\begin{aligned}
\sup_{s\in[\tau,t]} |\eta(s) - \xi(s)| &= \sup_{s\in[\tau,t]} \left| \frac{s-\tau}{t-\tau}x + \frac{t-s}{t-\tau}y - x + (t-s)H_p(t,x,p) + O((t-s)^2) \right| \\
&= \sup_{s\in[\tau,t]} \bigg| \frac{t-s}{t-\tau}\big(x - (t-\tau)H_p(t,x,p) + O((t-\tau)^2)\big) - \frac{t-s}{t-\tau}x \\
&\qquad + (t-s)H_p(t,x,p) + O((t-s)^2) \bigg| \\
&\leq O((t-\tau)^2).
\end{aligned}$$

Moreover from

$$y_p = \xi_p(\tau) = -(t-\tau)H_{pp}(t,x,p) + O((t-\tau)^2),$$

and from

$$\xi_p(s) = -(t-s)H_{pp}(t,x,p) + O((t-s)^2)$$

for $s \in [\tau, t]$, we obtain

$$\begin{aligned}
\sup_{s\in[\tau,t]} |\eta_p(s) - \xi_p(s)| &= \sup_{s\in[\tau,t]} \left| \frac{t-s}{t-\tau}y_p + (t-s)H_{pp}(t,x,p) + O((t-s)^2) \right| \\
&= \sup_{s\in[\tau,t]} \bigg| \frac{t-s}{t-\tau}\big(-(t-\tau)H_{pp}(t,x,p) + O((t-\tau)^2)\big) \\
&\qquad + (t-s)H_{pp}(t,x,p) + O((t-s)^2) \bigg| \\
&\leq O((t-\tau)^2).
\end{aligned}$$

In an analogous way, from

$$y_{pp} = \xi_{pp}(\tau) = -(t-\tau)H_{ppp}(t,x,p) + O((t-\tau)^2),$$

and from

$$\xi_{pp}(s) = -(t-s)H_{ppp}(t,x,p) + O((t-s)^2)$$

for $s \in [\tau, t]$, we obtain

$$\sup_{s\in[\tau,t]} |\eta_{pp}(s) - \xi_{pp}(s)| \le O((t-\tau)^2).$$

Observe now that

$$\dot{\eta}(s) = \frac{x-y}{t-\tau},$$

and

$$\dot{\xi}(s) = -H_p(t,x,p) + O(t-s),$$

hence

$$\begin{aligned}
\sup_{s\in[\tau,t]} |\dot{\eta}(s) - \dot{\xi}(s)| &= \sup_{s\in[\tau,t]} \left| \frac{x-y}{t-\tau} - H_p(t,x,p) + O(t-s) \right| \\
&= \sup_{s\in[\tau,t]} \Big| H_p(t,x,p) + O(t-\tau) - H_p(t,x,p) + O(t-s) \Big| \\
&\le O(t-\tau).
\end{aligned}$$

In the same way we obtain

$$\sup_{s\in[\tau,t]} |\dot{\eta}_p(s) - \dot{\xi}_p(s)| \le O(t-\tau),$$

and

$$\sup_{s\in[\tau,t]} |\dot{\eta}_{pp}(s) - \dot{\xi}_{pp}(s)| \le O(t-\tau).$$

□

Now, fix $x \in \mathbb{R}^n$ and a compact set $K \subset \mathbb{R}^n$. Call $\xi(\tau, K)$ the subset of $\mathbb{R}^n$ defined as

$$\xi(\tau, K) := \{\xi(\tau) |\ \xi \text{ is a solution of (3.4) with final conditions (3.7)}\}.$$

For any y in $\xi(\tau, K)$ consider the function

$$\phi(\tau, y, t, x) := \min \left\{ \int_\tau^t L(s, \xi(s), \dot{\xi}(s)) ds \ \middle|\ \xi \in [C^2([\tau, t])]^n,\ \xi(\tau) = y,\ \xi(t) = x \right\},$$

and observe that for any $y \in \xi(\tau, K)$ there exists a unique ξ solution of (3.4) with final conditions (3.7) such that $y = \xi(\tau, p)$. Thus we can see y as $y = y(p)$ with a C^2 dependence of y from p.

Proposition 3.13. *It holds*

$$\left\| \phi(\tau, y(p), t, x) - (t-\tau) L\left(t, x, \frac{x - y(p)}{t-\tau}\right) \right\|_{C^2(K)} \le O((t-\tau)^2).$$

In particular for $t-\tau$ small enough $y \mapsto \phi(\tau,y,t,x)$ and $x \mapsto \phi(\tau,y,t,x)$ are convex with constant $\frac{c_H}{t-\tau}$, in the sense that $\partial_{yy}\phi(\tau,y,t,x) \geq \frac{c_H}{t-\tau} Id_n$ and $\partial_{xx}\phi(\tau,y,t,x) \geq \frac{c_H}{t-\tau} Id_n$.

Proof. Note that, from the definition, $y \mapsto \phi(\tau,y,t,x)$ is automatically semiconcave.

Moreover, it is enough to consider the function $y \mapsto \phi(\tau,y,t,x)$ since there is a symmetry between the definitions of the functions $y \mapsto \phi(\tau,y,t,x)$ and $x \mapsto \phi(\tau,y,t,x)$.

From the definition, the function $y \mapsto \phi(\tau,y,t,x)$ has a unique minimum ξ which is the solution to system (3.4) with final conditions (3.7). Thus the C^2 dependence of y from p implies, for a small $t-\tau$, that $p \mapsto \phi(\tau,y(p),t,x)$ belongs to $C^2(K)$.

Let ξ be the unique minimizer for $\phi(\tau,y,t,x)$ and observe that $x=\eta(t)$ and $\frac{x-y}{t-\tau} = \dot\eta(t)$, where η is the straight line joining x to y as in (3.10).

$$\sup_{p\in K}\left|\phi(\tau,y(p),t,x)-(t-\tau)L\left(t,x,\frac{x-y(p)}{t-\tau}\right)\right| =$$

$$\begin{aligned}
&= \sup_{p\in K}\left|\int_\tau^t L(s,\xi(s),\dot\xi(s))ds - \int_\tau^t L(t,\eta(t),\dot\eta(t))ds\right| \\
&\leq \sup_{p\in K}\left\{\left|\int_\tau^t L(s,\xi(s),\dot\xi(s))ds - \int_\tau^t L(t,\xi(s),\dot\xi(s))ds\right|\right. \\
&\quad + \left|\int_\tau^t L(t,\xi(s),\dot\xi(s))ds - \int_\tau^t L(t,\eta(t),\dot\xi(s))ds\right| \\
&\quad \left.+ \left|\int_\tau^t L(t,\eta(t),\dot\xi(s))ds - \int_\tau^t L(t,\eta(t),\dot\eta(t))ds\right|\right\} \\
&\leq \sup_{p\in K}\left\{C_1\int_\tau^t |s-t|ds + C_2\int_\tau^t |\xi(s)-\eta(t)|ds + C_3\int_\tau^t |\dot\xi(s)-\dot\eta(t)|ds\right\} \\
&\leq \sup_{p\in K}\left\{-\frac{C_1}{2}(t-\tau)^2 + C_2 H_p(t,x,p)\int_\tau^t |t-s|ds\right. \\
&\quad \left.+ C_2\int_\tau^t O((t-s)^2)ds + C_3\int_\tau^t O((t-s))ds\right\} \\
&\leq O((t-s)^2).
\end{aligned}$$

Moreover for the first derivative

$$\sup_{p\in K}\left|\partial_p\left[\phi(\tau,y(p),t,x)-(t-\tau)L\left(t,x,\frac{x-y(p)}{t-\tau}\right)\right]\right| =$$

$$\begin{aligned}
&= \sup_{p\in K}\left|\int_\tau^t L_x(s,\xi(s),\dot\xi(s))\xi_p(s)ds + \int_\tau^t L_v(s,\xi(s),\dot\xi(s))\dot\xi_p(s)ds\right. \\
&\quad \left.- \int_\tau^t L_x(t,\eta(t),\dot\eta(t))\eta_p(t)ds - \int_\tau^t L_v(t,\eta(t),\dot\eta(t))\dot\eta_p(t)ds\right| \\
&\leq \sup_{p\in K}\left\{\left|\int_\tau^t L_x(s,\xi(s),\dot\xi(s))(-(t-s)H_p(t,x,p) + O(t-s)^2)ds\right|\right.
\end{aligned}$$

$$
\begin{aligned}
&+\left|\int_\tau^t L_v(s,\xi(s),\dot\xi(s))(-H_{pp}(t,x,p)+O(t-s))\right.\\
&\left.-L_v(t,\eta(t),\dot\eta(t))(-H_{pp}(t,x,p)+O(t-\tau))ds\right|\Bigg\}\\
\le{}& \sup_{p\in K}\Bigg\{C_1\int_\tau^t|(s-t)+O(t-s)|ds\\
&+\left|\int_\tau^t[L_v(s,\xi(s),\dot\xi(s))-L_v(t,\eta(t),\dot\eta(t))](C_2+O(t-\tau))ds\right|\Bigg\}\\
\le{}& O((t-s)^2).
\end{aligned}
$$

Analogously for the second derivative

$$
\sup_{p\in K}\left|\partial_{pp}\left[\phi(\tau,y(p),t,x)-(t-\tau)L\left(t,x,\frac{x-y(p)}{t-\tau}\right)\right]\right|\le O((t-s)^2).
$$

The map $p\mapsto y(p)$ is $C^2(K)$, it has bounded derivative and the same holds true also for its inverse, due to Proposition 3.12. Thus it follows that $\phi(\tau,y,t,x)$ and $(t-\tau)L\left(t,x,\frac{x-y}{t-\tau}\right)$ are close in $C^2(\tilde K)$, where $\tilde K$ is the image of K through the map $p\mapsto y(p)$. Therefore $y\mapsto\phi(\tau,y,t,x)$ is convex with constant $\frac{c_H}{t-\tau}$, the same constant of $y\mapsto(t-\tau)L\left(t,x,\frac{x-y}{t-\tau}\right)$. □

REMARK 3.14. *All the estimates found strictly depend on the compact set K, however thanks to the finite speed of propagation of the minimizers ξ, see point* (iii) *of Theorem 3.1, the estimates can be made uniform for our $\tilde u$.*

Let us now come back to our case.

PROPOSITION 3.15. *For $0\le\tau<t$ consider the backward solution defined in (3.6) for y in Ω_τ. Then for $t-\tau$ small enough the maximum is unique for all $y\in\Omega_\tau$.*

Proof. The backward solution can be written in this equivalent way

$$
\tilde u(\tau,y)=\max_{x\in\Omega_t}\{u(t,x)-\phi(\tau,y,t,x)\}. \tag{3.11}
$$

Recalling that $u(t,\cdot)$ is semiconcave with constant $\frac{C}{t}$ and that $-\phi(\tau,y,t,\cdot)$ is strictly concave with constant $\frac{c_H}{t-\tau}$, we can rewrite (3.11) as

$$
\tilde u(\tau,y)=\max_{x\in\Omega_t}\left\{u(t,x)-\frac{C}{t}|x|^2-\phi(\tau,y,t,x)+\frac{C}{t}|x|^2\right\}.
$$

Hence, since $u(t,x)-\frac{C}{t}|x|^2$ is concave and $-\phi(\tau,y,t,x)+\frac{C}{t}|x|^2$ remains strictly concave, the function $\tilde u(\tau,y)$ is the maximum of a strictly concave function, hence this maximum is unique. Thus there exists a unique $x\in\Omega_t$ such that

$$
\tilde u(\tau,y)=u(t,x)-\phi(\tau,y,t,x),
$$

i.e. there exists a unique curve $\xi\in[C^2([\tau,t])]^n$ such that $\xi(\tau)=y$, $\xi(t)=x$ and

$$
\tilde u(\tau,y)=u(t,\xi(t))-\int_\tau^t L(s,\xi(s),\dot\xi(s))ds.
$$

□

COROLLARY 3.16. *For $t-\tau$ small enough and $s \in (\tau, t)$ the function $\tilde{u}(s,\cdot)$ is $C^{1,1}(\Omega_s)$.*

Proof. From the above proposition we know that $\tilde{u}(s,\cdot)$ is $C^1(\Omega_s)$ for every $s \in [\tau, t)$. Consider now the *forward solution* defined from $\tilde{u}(\tau,\cdot)$

$$\hat{u}(s,x) := \min \left\{ \tilde{u}(\tau, \xi(\tau)) + \int_\tau^s L(l, \xi(l), \dot{\xi}(l))dl \;\Big|\; \xi(s) = x,\ \xi \in [C^2([\tau, s])]^n \right\}.$$

Due to the fact that $\tilde{u}(\tau, y)$ has a unique maximizer for every $y \in \Omega_\tau$ we have that $\hat{u}(s,x) = \tilde{u}(s,x)$ for every $s \in [\tau, t]$ and $x \in \Omega_s$. Thus for $s \in (\tau, t)$, $\tilde{u}(s,\cdot)$ is both semiconvex and semiconcave, hence $C^{1,1}(\Omega_s)$. □

REMARK 3.17. *As a consequence of Proposition 3.15, for every $y \in \Omega_\tau$ there exists only one curve which is a maximizer for the function $\tilde{u}(\tau, y)$ and a generalized backward characteristic. Hence generalized backward characteristics which are also maximizers do not intersect even at time τ. It remains to prove the following.*

PROPOSITION 3.18. *Every generalized backward characteristic $\xi(\cdot)$, i.e. a solution of (3.4) with final conditions (3.5) where $p \in D_x^+ u(t,x)$, is a maximizer for $\tilde{u}(\tau, \xi(\tau))$ if $t-\tau$ is small enough.*

Proof. Let ξ be a generalized backward characteristic with $\xi(t) = x$, $p(t) = p \in D_x^+ u(t,x)$ and $\xi(\tau) = y$. Then ξ is a minimizer for $\phi(\tau, t, y, x)$ and $p = p(t) = -D_y \phi(\tau, y, t, x)$.

Let $\tilde{\xi}$ be the unique maximizer for $\tilde{u}(\tau, y)$ and suppose by contradiction that $\tilde{\xi}$ differs from ξ, in particular $\tilde{\xi}(t) = \tilde{x} \neq x = \xi(t)$. Then by definition

$$\tilde{u}(\tau, y) = u(t, \tilde{x}) - \phi(\tau, y, t, \tilde{x}) > u(t,x) - \phi(\tau, y, t, x).$$

Thus, for the differentiability and the convexity of $\phi(\tau, y, t, \cdot)$

$$\begin{aligned} u(t,\tilde{x}) - u(t,x) &> \phi(\tau, y, t, \tilde{x}) - \phi(\tau, y, t, x) \\ &\geq \langle D_y \phi(\tau, y, t, x), \tilde{x} - x \rangle + \frac{c_H}{t-\tau} |\tilde{x} - x|^2. \end{aligned}$$

On the other hand for the semiconcavity of $u(t,\cdot)$

$$u(t,\tilde{x}) - u(t,x) < \langle p, \tilde{x} - x \rangle + \frac{C}{t} |\tilde{x} - x|^2.$$

Thus, recalling that $p = -D_y \phi(\tau, y, t, x)$, for $t - \tau$ small enough we reach the absurd

$$\frac{C}{t} > \frac{c_H}{t-\tau}.$$

□

From the above proposition it follows

COROLLARY 3.19. *Generalized backward characteristics cannot intersect in $[\tau, t)$ if $t-\tau$ is small enough.*

3.2. Local property. Thanks to the time invariance of the equation and to the following locality property, which is a generalization of the Proposition 3.5 found in [5], it is enough to prove Theorem 1.1 for the unique viscosity solution of (1.1) with a Lipschitz bounded initial datum

$$u(0,\cdot) = u_0(\cdot). \tag{3.12}$$

PROPOSITION 3.20. *Let u be a viscosity solution of (1.1) in Ω. Then u is locally Lipschitz. Moreover for any $(t_0, x_0) \in \Omega$, there exists a neighborhood $\mathcal{U}$ of (t_0, x_0), a positive number δ and a Lipschitz function v_0 on $\mathbb{R}^n$ such that u coincides on $\mathcal{U}$ with the viscosity solution of*

$$\begin{cases} \partial_t v + H(t, x, D_x v) = 0 & \text{in } [t_0 - \delta, \infty) \times \mathbb{R}^n \\ v(t_0 - \delta, x) = v_0(x). \end{cases}$$

Proof. The proof of Proposition 3.5, given in [5], still applies in our case where we only lose the property that minimizers of (3.1) are straight lines which was unnecessary for the argument. □

4. Proof of the main theorem.

4.1. Preliminary remarks. Let u be a viscosity solution of (1.1). Applying Proposition 3.20 we can assume without loss of generality that u is a solution of the Cauchy Problem (1.1)-(3.12) over a bounded domain $[0, \delta] \times U$ and with a bounded and Lipschitz initial datum. Moreover assumptions (H1)-(H2) guarantee that the Hamiltonian is convex and has super-linear growth in the last variable.

We will prove the SBV regularity over the smaller interval of time $[\tau, \tau + \varepsilon]$ for a fixed $\tau > 0$, $\varepsilon > 0$ small enough and such that $[\tau, \tau + \varepsilon] \subset [0, \delta]$. As we have already seen, this is necessary to prevent intersections of generalized backward characteristics.

We consider a ball $B_R(0) \subset \mathbb{R}^n$ and a bounded convex set $\Omega \subset [\tau, \tau + \varepsilon] \times \mathbb{R}^n$ with the properties that

- $\{s\} \times B_R(0) \subset \Omega$ for every $s \in [\tau, \tau + \varepsilon]$;
- for any $(t, x) \in \Omega$ and for any C^2 curve ξ which minimizes $u(t, x)$ in (3.1), the entire curve $\xi(s)$ for $s \in [\tau, t]$ is contained in Ω.

Indeed, from the fact that $\|Du\|_\infty < \infty$, it is enough to choose

$$\Omega := \{(t, x) \in [\tau, \tau + \varepsilon] \times \mathbb{R}^n |\ |x| \leq R + C'(\tau + \varepsilon - t)\}$$

with C' sufficiently large and depending only on $\|Du\|_\infty$ and H.

The general idea of the proof is now standard, see [2], [5]. We construct a monotone bounded functional $F(t)$ defined on the interval $[\tau, \tau + \varepsilon]$. Then, we relate the presence of a Cantor part in the matrix $D_x^2 u(t, \cdot)$ for a certain t in $[\tau, \tau + \varepsilon]$ with a jump of the functional F in t. Since this functional can have only a countable number of jumps, the Cantor part of $D_x^2 u(t, \cdot)$ can be different from zero only for a countable number of t's.

REMARK 4.1. *Once we have formalized the above strategy and proved the SBV regularity for almost every t in $[\tau, \tau + \varepsilon]$ the conclusion that $D_x u$ belongs to $[SBV_{loc}(\Omega)]^n$ follows from the slicing theory of BV functions (see Theorem* 3.108 *of [3]). The local SBV regularity of $\partial_t u$ follows instead from the Volpert chain rule.*

4.2. Construction of the functional F. Consider t belonging to $(\tau, \tau + \varepsilon]$ for a fixed $\tau > 0$ and $\varepsilon > 0$ small enough. For any $\tau \leq s < t$ we define the set-valued map

$$X_{t,s}(x) := \{\xi(s) |\ \xi(\cdot) \text{ is a solution of (3.4), with } \xi(t) = x,\ p(t) = p \in D_x^+ u(t, x)\}.$$

Moreover we will denote by $\chi_{t,s}$ the restriction of $X_{t,s}$ to the points where it is single-valued. According to Theorem 3.6, the domain of $\chi_{t,s}$, $dom(\chi_{t,s}) =: U_t$, consists of those points where $D_x^+ u(t, x)$ is single-valued, i.e. there exists a unique minimizer for

$u(t,x)$ in the representation formula (3.1). For that reason $\chi_{t,s}$ is clearly defined a.e. in Ω_t. We will sometimes write $\chi_{t,s}(\Omega_t)$ meaning $\chi_{t,s}(U_t)$.

REMARK 4.2. *In the definition of $X_{t,s}$ we follow generalized backward characteristics starting at time $t>0$ till time s. As we have already seen, if $t-s$ is small enough, generalized backward characteristics cannot intersect except than at time t. Thus if we choose $\varepsilon>0$ small enough we have the injectivity of the set valued map $X_{t,\tau}$ over the interval of time $[\tau,\tau+\varepsilon]$.*

Note that in the case $H=H(D_xu)$ the authors of [5] were able, in Proposition 5.2*, to prove the injectivity of $X_{t,0}$, as a set-valued map, for every $t\in[0,\varepsilon]$ with ε small enough.*

Therefore, equivalently to Proposition 5.2 in [5], we can state

PROPOSITION 4.3. *Let t be fixed such that $\tau<t\leq\tau+\varepsilon$, for an $\varepsilon>0$ small enough, which does not depend on t. Then taken any two solutions (ξ_1,p_1) and (ξ_2,p_2) of the system (3.4) with final condition*

$$\xi_i(t)=x_i\in\Omega_t \qquad p_i(t)\in D_x^+u(t,x_i) \qquad i=1,2,$$

and $(\xi_1(t),p_1(t))\neq(\xi_2(t),p_2(t))$ it follows that $\xi_1(\tau)\neq\xi_2(\tau)$. Hence, in particular, the map $x\mapsto X_{t,\tau}(x)$ is injective as a set-valued map.

Proof. It follows from Corollary 3.19. □

For every $\tau<t\leq\tau+\varepsilon$, we can now define the functional

$$F(t):=\mathcal{H}^n(\chi_{t,\tau}(U_t)). \tag{4.1}$$

LEMMA 4.4. *The functional F is non increasing,*

$$F(s)\geq F(t) \quad \text{for any } s,t\in(\tau,\tau+\varepsilon] \text{ with } s<t.$$

Proof. As in the proof of Lemma 4.1 in [5], the claim follows from the following consideration:

$$\chi_{t,\tau}(\Omega_t)\subset\chi_{s,\tau}(\Omega_s) \quad \text{for every } \tau\leq s\leq t\leq\tau+\varepsilon.$$

Indeed, consider any $y\in\chi_{t,\tau}(\Omega_t)$. Then there exists a $[C^2([\tau,t])]^n$ curve ξ and a point $x\in\Omega_t$ such that ξ is the unique minimizer in (3.1) with the following endpoints conditions $\xi(t)=x,\xi(\tau)=y$. Such a curve remains the unique minimizer also for $u(s,\xi(s))$ for any $\tau\leq s\leq t\leq\tau+\varepsilon$. Hence, setting $z=\xi(s)$, we have that the point y can be seen as $y=\chi_{s,\tau}(z)$ and $y\in\chi_{s,\tau}(\Omega_s)$. □

4.3. Hille-Yosida transformation. Take a Borel set $A\subset\Omega_t$ for a fixed time $t\in(\tau,\tau+\varepsilon]$. In order to compute the measure $\mathcal{H}^n(X_{t,\tau}(A))$ we follow the evolution of the set along generalized backward characteristics till the time τ.

Let us recall how the characteristics and their dual arc evolve in time. They are solutions of the system (3.4), together with the final condition (3.5) where p belongs to $D_x^+u(t,x)$.

We have to face the following problem: the function $D_x^+u(t,\cdot)$ is a multi-valued function which is not Lipschitz in general. However it can be easily related to a maximal monotone function whose graph can be parametrized in a Lipschitz way as shown in Alberti and Ambrosio [1].

Let us consider the graph $(A,D_x^+u(t,A))$ for a Borel set $A\subset\Omega_t$. Since $u(t,x)$ is semiconcave in x, $v(x):=-(u(t,x)-\frac{1}{2}C|x|^2)$ is a convex function. Note that the semiconcavity constant should depend on t, i.e. $C(t) = \frac{C}{t}$, however a uniform one can be taken due to the fact that t belongs to $(\tau, \tau + \varepsilon]$ where $\tau > 0$. Moreover, as seen in Theorem 2.5-(iv), the superdifferential of v is a maximal monotone function. It can be proven, see for example [1], that the graph of a maximal monotone function is a Lipschitz submanifold without boundary. Adapting the same procedure to our case, we can parametrize the graph of the gradient of our semiconcave function with a 1-Lipschitz function.

Indeed, we pass from our graph $\{(x, D_x^+ u(t,x)) |\, x \in A\}$ to the graph of a maximal monotone function with the following transformation

$$\begin{cases} x = x \\ y = Cx - p, \end{cases}$$

where C is the semiconcavity constant of $u(t, \cdot)$. Then we apply an Hille-Yosida transformation to have a 1-Lipschitz parametrization of it.

$$\begin{cases} z = x + y \\ w = y. \end{cases}$$

Call $T(x) := D_x v(x)$ the maximal monotone function. Retracing the passages above, we can express w as a 1-Lipschitz single-valued function of z. Taking $z \in B := A + T(A)$

$$\begin{cases} z = z \\ w = (Id_n + (T)^{-1})^{-1}(z). \end{cases} \tag{4.2}$$

Thus, coming back to our original coordinates, we can describe our graph with the following Lipschitz parametrization

$$\begin{cases} x(z) = z - w(z) \\ p(z) = Cz - (C+1)w(z), \end{cases} \tag{4.3}$$

where $z \in B$, i.e. we have

$$\Gamma_A := \{(x, D_x^+ u(t,x)) |\, x \in A\} = \{(z - w(z), Cz - (C+1)w(z)) |\, z \in B\}.$$

REMARK 4.5. *As explained in [1] the* 1*-Lipschitz function* $w(\cdot)$ *is exactly the gradient of the inf-convolution function of* $v(x) = -(u(t,x) - \frac{1}{2}C|x|^2)$

$$f(z) = \min_{x \in \mathbb{R}^n} \left\{ v(x) + \frac{|x - z|^2}{2} \right\}.$$

Thus we have $w(z) = f_z(z)$ *where* f *is the convex function above.*

When applying the flux backward in time, starting from our set Γ_A, characteristics $\xi(\cdot, z)$ and $p(\cdot, z)$ evolve according to

$$\begin{cases} \dot{\xi}(s,z) = H_p(s, \xi(s,z), p(s,z)) \\ \dot{p}(s,z) = -H_x(s, \xi(s,z), p(s,z)) \end{cases} \tag{4.4}$$

with final conditions

$$\begin{cases} \xi(t,z) = x(z) = z - w(z) \\ p(t,z) = p(z) = Cz - (C+1)w(z), \end{cases} \tag{4.5}$$

for z in B. Since the flux is described by smooth equations and thanks to the fact that the parametrization of our initial set is 1-Lipschitz, the solutions $\xi(\cdot,z), p(\cdot,z)$ are Lipschitz curves.

We can now rewrite $X_{t,\tau}$ in an equivalent way, for x in A

$$\begin{aligned} X_{t,\tau}(x) &= \{\xi(\tau)\ |\xi(\cdot) \text{ is a solution of (3.4), with } \xi(t)=x,\ p(t)=p\in D_x^+u(t,x)\} \\ &= \{\xi(\tau,z)\ |\xi(\cdot,z) \text{ is a solution of (4.4), with } \xi(t,z)=z-w(z), \\ &\quad p(t,z)=Cz-(C+1)w(z), z\in x+T(x)\}. \end{aligned}$$

With an abuse of notation we will denote with $\xi(\tau,\cdot): B\to\Omega_\tau$ the function $X_{t,\tau}(\cdot)$ when we are considering the Lipschitz parametrization; with this notation $X_{t,\tau}(A)=\xi(\tau,B)$. We can now apply the Area Formula to $\xi(\tau,\cdot)$

$$\int_{\xi(\tau,B)}\mathcal{H}^0((\xi(\tau,\cdot)^{-1}(w))dw=\int_B|\det(\xi_z(\tau,z))|dz. \tag{4.6}$$

Thanks to the injectivity of the map $X_{t,\tau}$ which is preserved when passing to the Lipschitz parametrization, the left term of (4.6) is precisely the measure of the set $\xi(\tau,B)$.

Hence, we have

$$\int_{\xi(\tau,B)}\mathcal{H}^0((\xi(\tau,\cdot)^{-1}(w))dw=\mathcal{H}^n(\xi(\tau,B))=\mathcal{H}^n(X_{t,\tau}(A)).$$

To compute $\det(\xi_z(\tau,z))$ we differentiate in z the equations (4.4), (4.5) obtaining that ξ_z and p_z satisfy the system

$$\begin{cases} \dot{\xi}_z(s,z)=H_{px}(s,\xi(s,z),p(s,z))\xi_z(s,z)+H_{pp}(s,\xi(s,z),p(s,z))p_z(s,z) \\ \dot{p}_z(s,z)=-H_{xx}(s,\xi(s,z),p(s,z))\xi_z(s,z)-H_{xp}(s,\xi(s,z),p(s,z))p_z(s,z) \end{cases} \tag{4.7}$$

with the final conditions

$$\begin{cases} \xi_z(t,z)=Id_n(z)-w_z(z) \\ p_z(t,z)=CId_n(z)-(C+1)w_z(z), \end{cases} \tag{4.8}$$

for any $z\in B$.

4.4. Approximation. If we choose $\varepsilon>0$ small enough we can approximate our curves with straight lines for any t in $(\tau,\tau+\varepsilon]$, i.e. we can write

$$\xi(\tau,z)=\xi(t,z)-(t-\tau)\dot{\xi}(t,z)+O((t-\tau)^2).$$

Using this approximation and (4.7) we obtain

$$\begin{aligned} \det(\xi_z(\tau,z)) = {} & \det\Big(\xi_z(t,z)-(t-\tau)H_{px}(t,x(z),p(z))\xi_z(t,z) \\ & -(t-\tau)H_{pp}(t,x(z),p(z))p_z(t,z)\Big)+O((t-\tau)^2). \end{aligned} \tag{4.9}$$

Since we are now considering nearly straight lines, instead of more general curves, we can expect that this approximation should allow us to adapt the techniques of [5] and recover the lemmas needed.

Before going on, let us give an explicit formula for the spatial-Laplacian of our solution. Thanks to the semiconcavity of $u(t,\cdot)$ its spatial-Laplacian is a measure.

Moreover, using the 1-Lipschitz parametrization given by Hille-Yosida, the spatial-Laplacian can be seen as the push-forward of a particular measure.

LEMMA 4.6. *For any Borel set A, let $\{(x(z), p(z))|\ z \in A + T(A)\}$ be the 1-Lipschitz parametrization of the set $\{(x, D_x^+ u(t, x))|\ x \in A\}$ as seen above in (4.3). Then we have*

$$\Delta u(t, A) = x(z)_\sharp \left[\left(\sum_{i,k} \frac{\partial p_i(z)}{\partial z_k} [\text{cof } x_z(z)]_{ik} \right) \mathcal{H}^n \right] (A).$$

Here cof A *is the cofactor matrix of the matrix A.* This formula has been shown to the authors by C. De Lellis.

Proof. Take any ϕ in $C_c^\infty(\mathbb{R}^n)$ and compute

$$\begin{aligned}
\int_{\mathbb{R}^n} \phi(x) d[D_x^2 u(t,x)]_{ij} &= -\int_{\mathbb{R}^n} [D_x u(t,x)]_i \frac{\partial \phi(x)}{\partial x_j} dx \\
&= -\int_{\mathbb{R}^n} p_i(z) \frac{\partial \phi(x(z))}{\partial x_j} \det(x_z(z)) dz \\
&= -\int_{\mathbb{R}^n} p_i(z) \sum_k \left(\frac{\partial \phi(x(z))}{\partial z_k} \frac{\partial z_k(x(z))}{\partial x_j} \right) \det(x_z(z)) dz \\
&= -\int_{\mathbb{R}^n} p_i(z) \sum_k \left(\frac{\partial \phi(x(z))}{\partial z_k} [\text{cof } x_z(z)]_{jk} \right) dz \\
&= \int_{\mathbb{R}^n} \phi(x(z)) \sum_k \left(\frac{\partial p_i(z)}{\partial z_k} [\text{cof } x_z(z)]_{jk} \right) dz \\
&\quad - \int_{\mathbb{R}^n} \sum_k \left(\frac{\partial}{\partial z_k} \phi(x(z)) p_i(z) \right) [\text{cof } x_z(z)]_{jk} dz.
\end{aligned}$$

In the lines above we have used the 1-Lipschitz parametrization (4.3) of the set $\{(x, D_x^+ u(t, x))|\ x \in A\}$ and the fact that

$$\frac{\partial z_k(x(z))}{\partial x_j} = [x_z(z)]_{kj}^{-1} = \frac{1}{\det(x_z(z))} [\text{cof } x_z(z)]_{jk}.$$

Now, if T^ε is a maximal monotone smooth function with strictly positive symmetric derivative, such that its inverse $(T^\varepsilon)^{-1}$ has derivative $\geq \varepsilon Id_n$ and converges in $W^{1,1}$ to T^{-1}: denote with w^ε the corresponding $(1-\varepsilon)$-Lipschitz function converging in $W^{1,1}$ to w (4.2), and with $x^\varepsilon(\cdot)$, $p^\varepsilon(\cdot)$ the corresponding functions given by (4.3).

Repeating upside down the passages starting from the last term with the function T^ε, one obtains that

$$\int_{\mathbb{R}^n} \sum_k \left(\frac{\partial}{\partial z_k} \phi(x^\varepsilon(z)) p_i^\varepsilon(z) \right) [\text{cof } x_z^\varepsilon(z)]_{jk} dz = -\int_{\mathbb{R}^n} \frac{\partial}{\partial x_j} \big(\phi(x) [D_x u^\varepsilon(t,x)]_i \big) dx$$

which is equal to zero due to the fact that ϕ has compact support. Because of the finite speed of propagation, we can assume that the supports of $\phi(x^\varepsilon)$ are uniformly bounded, and thus by passing to the limit we conclude that the last term of (refecomputlapla) vanishes. □

We are now able to prove an analogous of Lemma 4.3 in [5].

LEMMA 4.7. *For ε small enough (depending only on the bound M for $\|H_{px}\|$), let $t \in (\tau, \tau+\varepsilon]$ and $A \subset \Omega_t$ be a Borel set. Then*

$$\mathcal{H}^n(X_{t,\tau}(A)) \geq C_1 \mathcal{H}^n(A) - C_2(t-\tau)\int_A d\Delta u(t,\cdot) + O((t-\tau)^2),$$

where C_1, C_2 are positive constants (depending on C, c_H). $\Delta u(t,\cdot)$ is the spatial-Laplacian of $u(t,\cdot)$.

Proof. Let us start from (4.9).

For $t-\tau$ small enough the matrix

$$Id_n(z) - (t-\tau)H_{px}(t, x(z), p(z))$$

is invertible. Indeed, since $\exists M > 0$ such that the norm $\|H_{px}(\cdot,\cdot,\cdot)\| < M$ it is sufficient to take $\varepsilon < \frac{1}{2nM}$. This condition ensures that

$$\det(Id_n(z) - (t-\tau)H_{px}(t, x(z), p(z))) > \frac{1}{2} > 0.$$

Thus this determinant can be picked out in (4.9)

$$\begin{aligned} |\det(\xi_z(\tau,z))| &= |\det(Id_n - (t-\tau)H_{px})|\,|\det\left(\xi_z - (t-\tau)(Id_n - (t-\tau)H_{px})^{-1}H_{pp}p_z\right)| \\ &\quad + O((t-\tau)^2) \\ &> \frac{1}{2}|\det(\xi_z - (t-\tau)H_{pp}p_z)| + O((t-\tau)^2). \end{aligned}$$

To lighten the computation above we have omitted the dependence of H_{px}, H_{pp} from $t, x(z), p(z)$ and of ξ_z, p_z from t, z. Moreover we used the fact that for $t-\tau$ small enough it is possible to expand the inverse

$$(Id_n - (t-\tau)H_{px})^{-1} = Id_n + (t-\tau)H_{px} + O((t-\tau)^2).$$

We are then left to expand the determinant in series

$$\det(\xi_z - (t-\tau)H_{pp}p_z) = \det(\xi_z) - (t-\tau)\mathrm{tr}\ \left([\mathrm{cof}\ \xi_z]^T H_{pp}p_z\right) + O((t-\tau)^2),$$

and use that $w = f_z$ as underlined in the Remark 4.5, so that, recalling (4.8),

$$\xi_z = Id_n - w_z = Id_n - f_{zz}, \qquad p_z = CId_n - (C+1)w_z = CId_n - (C+1)f_{zz}.$$

Call λ_i, for $i = 1, ..., n$, the eigenvalues of the positive semidefinite matrix f_{zz}. Hence we can compute

$$\det(\xi_z) = \prod_i (1-\lambda_i) \qquad [\mathrm{cof}\ \xi_z]_{ii} = \prod_{j\neq i}(1-\lambda_j).$$

The convexity of f and the 1-Lipschitzianity of f_z imply that all the eigenvalues are bounded from above and from below: $0 \leq \lambda_i \leq 1$, for $i = 1, \ldots, n$. Thus, for every $i = 1, \ldots, n$, we have $0 \leq 1 - \lambda_i \leq 1$ and $-1 \leq C - (C+1)\lambda_i \leq C$, in particular this last inequality suggests that we have to work a bit to bound our determinant, since $C - (C+1)\lambda_i$ has no definite sign.

$$\frac{1}{2}\left(\det(\xi_z) - (t-\tau)\mathrm{tr}\ \left([\mathrm{cof}\ \xi_z]^T H_{pp}p_z\right)\right) + O((t-\tau)^2) =$$

$$
\begin{aligned}
&= \frac{1}{2}\left(\prod_i (1-\lambda_i) - (t-\tau)\mathrm{tr}\ \left(\mathrm{diag}\left[\prod_{j\neq i}(1-\lambda_j)\right] H_{pp}\,\mathrm{diag}[C-(C+1)\lambda_i]\right)\right) \\
&\quad +O((t-\tau)^2) \\
&= \frac{1}{2}\left(\prod_i (1-\lambda_i) - (t-\tau)\sum_i\prod_{j\neq i}(1-\lambda_j)[H_{pp}]_{ii}(C-(C+1)\lambda_i)\right) + O((t-\tau)^2) \\
&= \frac{1}{2}\prod_i (1-\lambda_i) - (t-\tau)\frac{1}{2}\sum_i\prod_{j\neq i}(1-\lambda_j)[H_{pp}]_{ii}(C(1-\lambda_i)-\lambda_i) + O((t-\tau)^2) \\
&= \frac{1}{2}\left(1-(t-\tau)C\ \mathrm{tr}\ H_{pp}\right)\prod_i(1-\lambda_i) + (t-\tau)\frac{1}{2}\sum_i\lambda_i[H_{pp}]_{ii}\prod_{j\neq i}(1-\lambda_j) \\
&\quad +O((t-\tau)^2).
\end{aligned}
$$

Now that all the terms have positive sign for an ε small enough, we can use the uniform convexity of H in p and the bounds on λ_i to show that there exist constants C_1, C_2, all of them depending only on C, c_H, such that

$$
\begin{aligned}
|\det(\xi_z(\tau,z))| &\geq C_1\prod_i(1-\lambda_i) + (t-\tau)C_2\sum_i\lambda_i\prod_{j\neq i}(1-\lambda_j) + O((t-\tau)^2) \\
&\geq C_1\prod_i(1-\lambda_i) + (t-\tau)C_2\sum_i\lambda_i\prod_{j\neq i}(1-\lambda_j) \\
&\quad -n(t-\tau)C_2C\prod_i(1-\lambda_j) + O((t-\tau)^2) \\
&= C_1\prod_i(1-\lambda_i) - (t-\tau)C_2\sum_i(C(1-\lambda_i)-\lambda_i))\prod_{j\neq i}(1-\lambda_j) + O((t-\tau)^2) \\
&= C_1\prod_i(1-\lambda_i) - (t-\tau)C_2\sum_i(C-(C+1)\lambda_i)\prod_{j\neq i}(1-\lambda_j) + O((t-\tau)^2).
\end{aligned}
$$

Therefore if we compute the area formula (4.6) we obtain

$$
\int_B |\det(\xi_z(\tau,z))|dz \geq
$$

$$
\geq \int_B\left[C_1\prod_i(1-\lambda_i) - (t-\tau)C_2\sum_i(C-(C+1)\lambda_i)\prod_{j\neq i}(1-\lambda_j)\right]dz + O((t-\tau)^2).
$$

Applying Lemma 4.6 and recalling that $1-\lambda_i$ are the eigenvalues of $\xi_z(t,z)$ we obtain the thesis.

$$
\mathcal{H}^n(X_{t,\tau}(A)) \geq C_1\mathcal{H}^n(A) - C_2(t-\tau)\int_A d\Delta u(t,\cdot) + O((t-\tau)^2)
$$

where C_1, C_2 are constants depending only on C, c_H. □

4.5. Area estimates. In order to complete the proof of the main theorem we need to prove a Lemma which states the equivalent result of Lemma 5.1 in [5].

LEMMA 4.8. *If $\varepsilon > 0$ is small enough, for any $t \in (\tau, \tau+\varepsilon]$, any $\delta \in [0, t-\tau]$ and any Borel set $A \subset \Omega_t$ we have*

$$
\mathcal{H}^n(X_{t,\tau+\delta}(A)) \geq \left(\frac{1}{2}\right)^n\left(\frac{t-(\tau+\delta)}{t-\tau}\right)^n\mathcal{H}^n(X_{t,\tau}(A)).
$$

Proof. Fix t in $(\tau, \tau+\varepsilon]$, and let A be a Borel set $A \subset \Omega_t$. Without loss of generality we can suppose A to be a compact set.

Consider an approximation of the vector field induced by our generalized backward characteristics by taking a dense sequence of points $\{x_i\}_{i=1}^{\infty}$ in A. Fix an integer $I > 0$, call $A_I := \{x_i |\ i = 1, \dots, I\}$ and define for any s such that $\tau \le s < t$ and $y \in X_{t,s}(A)$

$$\tilde{u}_I(s,y) := \max \Bigg\{ u(t,\xi(t)) - \int_s^t L(l,\xi(l),\dot{\xi}(l))dl \ \Bigg|\ \xi \text{ is a } [C^2([s,t])]^n \text{ curve}, \\ \xi(s) = y,\ \xi(t) \in A_I \Bigg\}.$$

We assume in addition that the sequence $\{x_i\}_{i\in I}$ is big enough so that we can uniformly bound the speed of propagation of every maximizer ξ.

Remark 4.9. *All the properties which we stated for maximizers of the backward solution and for the backward solution itself are preserved in each cone of propagation for the maximizers of this approximated backward solution (Euler equation, systems for maximizer and dual arc, no-crossing property, etc) and for $\tilde{u}_I$ (a.e. differentiability, dynamic programming principle, semiconvexity).*

Through this approximation the set $E_s := X_{t,s}(A)$ is split into at most I open regions E_s^i, $i = 1, \dots, I$, defined by

$$E_s^i := \text{interior of } \{y \in X_{t,s}(A) |\ \exists \xi \text{ maximizer for } \tilde{u}_I(s,y) \text{ such that } \xi(t) = x_i\},$$

together with the set

$$J_s^I := \bigcup_{i \neq j} \big(\bar{E}_s^i \cap \bar{E}_s^j\big)$$

of negligible $\mathcal{H}^n$-measure. Indeed, even for $\tilde{u}_I(s,\cdot)$ the set of points with more than one maximum is the set of point of non differentiability and this set has $\mathcal{H}^n$-measure zero.

Call

$$X_{t,s}^I(x_i) := \{\xi(s) |\ \xi \text{ is a maximizer for } \tilde{u}_I(s,y) \text{ with } y \in \bar{E}_s^i\},$$

this is a multi-valued function defined on the set A_I.

The set $X_{t,s}^I(A_I)$ converges in the Hausdorff sense to the set $X_{t,s}(A)$ as I tends to infinity. Indeed, it follows from the strong convergence of the maximizers of $\tilde{u}^I$ to the maximizers of $\tilde{u}$ which is ensured by their bound on the derivative (Theorem 3.1-(iii)). Thus

$$\mathcal{H}^n(X_{t,s}(A)) \ge \limsup_{I\to\infty} \mathcal{H}^n(X_{t,s}^I(A_I)).$$

Let us decompose $\mathcal{H}^n(X_{t,s}^I(A_I))$ in the sum over $i \in I$ of $\mathcal{H}^n(X_{t,s}^I(x_i))$. Using the one to one correspondence of Lemma 3.11

$$\frac{\xi_p(\tau)}{\tau - t} = H_{pp}(t, x_i, p) + O(t-\tau)$$

and

$$\frac{\xi_p(\tau+\delta)}{\tau+\delta-t} = H_{pp}(t,x_i,p) + O(t-\tau).$$

Therefore

$$\left|\frac{\xi_p(\tau)}{\tau-t} - \frac{\xi_p(\tau+\delta)}{\tau+\delta-t}\right| \leq O(t-\tau),$$

and

$$\left|\left(\frac{t-(\tau+\delta)}{t-\tau}\right)\xi_p(\tau)(\xi_p(\tau+\delta))^{-1} - Id\right| \leq O(t-\tau).$$

Thus, passing to the determinant,

$$\det(\xi_p(\tau+\delta)) \geq \left(\frac{1}{2}\right)^n \left(\frac{t-(\tau+\delta)}{t-\tau}\right)^n \det(\xi_p(\tau)).$$

From which it follows

$$\mathcal{H}^n(X^I_{t,\tau+\delta}(x_i)) \geq \left(\frac{1}{2}\right)^n \left(\frac{t-(\tau+\delta)}{t-\tau}\right)^n \mathcal{H}^n(X^I_{t,\tau}(x_i)).$$

Summing up all the terms

$$\mathcal{H}^n(X^I_{t,\tau+\delta}(A_I)) \geq \left(\frac{1}{2}\right)^n \left(\frac{t-(\tau+\delta)}{t-\tau}\right)^n \mathcal{H}^n(X^I_{t,\tau}(A_I)).$$

Finally using the fact that $\mathcal{H}^n(X^I_{t,\tau}(A_I)) = \mathcal{H}^n(X_{t,\tau}(A))$ and the Hausdorff convergence we obtain

$$\begin{aligned}
\mathcal{H}^n(X_{t,\tau+\delta}(A)) &\geq \limsup_{I\to\infty} \mathcal{H}^n(X^I_{t,\tau+\delta}(A_I)) \\
&\geq \limsup_{I\to\infty} \left(\frac{1}{2}\right)^n \left(\frac{t-(\tau+\delta)}{t-\tau}\right)^n \mathcal{H}^n(X^I_{t,\tau}(A_I)) \\
&= \left(\frac{1}{2}\right)^n \left(\frac{t-(\tau+\delta)}{t-\tau}\right)^n \mathcal{H}^n(X_{t,\tau}(A)).
\end{aligned}$$

Hence the thesis is proved. □

We can now prove the following Lemma. In the following we will denote the Cantor part of $D^2_x u(t,\cdot)$ with $D^2_c u(t,\cdot)$.

LEMMA 4.10. *For ε small enough, for any t in $(\tau,\tau+\varepsilon]$ such that $|D^2_c u(t,\cdot)|(\Omega_t) > 0$ there exists a Borel set $A \subset \Omega_t$ such that*

i) $\mathcal{H}^n(A) = 0$, $|D^2_c u(t,\cdot)|(A) > 0$ and $|D^2_c u(t,\cdot)|(\Omega_t \setminus A) = 0$;
ii) $X_{t,\tau}$ is single-valued on A;
iii) and for any δ in $(0,\tau+\varepsilon-t]$,

$$\chi_{t,\tau}(A) \cap \chi_{t+\delta,\tau}(\Omega_{t+\delta}) = \emptyset.$$

Proof. From Proposition 2.6 and the definition of Cantor part of a measure, there exists a Borel set A such that

- $D^+_x u(t,x)$ is single-valued for every $x \in A$,

- $\mathcal{H}^n(A) = 0$,
- $|D_c^2 u(t,\cdot)|(\Omega_t \setminus A) = 0$ and $|D_c^2 u(t,\cdot)|(A) > 0$.

By contradiction suppose there exists a compact set $K \subset A$ such that

$$|D_c^2 u(t,\cdot)|(K) > 0$$

and there exists a $\delta > 0$ such that

$$X_{t,\tau}(K) = \chi_{t,\tau}(K) \subset \chi_{t+\delta,\tau}(\Omega_{t+\delta}).$$

Call $\omega := |D_c^2 u(t,\cdot)|(K)$.

Then there exists a Borel set $\tilde{K} \subset \Omega_{t+\delta}$ such that $\chi_{t,\tau}(K) = \chi_{t+\delta,\tau}(\tilde{K})$. Moreover, thanks to the fact that we are considering classical characteristics starting from $\tilde{K}$, we have

$$\chi_{t+\delta,t}(\tilde{K}) = K \quad \text{and} \quad \chi_{t+\delta,s}(\tilde{K}) = \chi_{t,s}(K) \quad \forall s \in [\tau, t).$$

Using Lemma 4.8, for any $s \in [\tau, t)$,

$$\begin{aligned}\mathcal{H}^n(K) = \mathcal{H}^n(X_{t+\delta,t}(\tilde{K})) &\geq \left(\frac{1}{2}\right)^n \left(\frac{\delta}{t+\delta-s}\right)^n \mathcal{H}^n(X_{t+\delta,s}(\tilde{K})) \\ &= \left(\frac{1}{2}\right)^n \left(\frac{\delta}{t+\delta-s}\right)^n \mathcal{H}^n(X_{t,s}(K)).\end{aligned}$$

Hence

$$\mathcal{H}^n(K) \geq \left(\frac{1}{2}\right)^n \left(\frac{\delta}{t+\delta-s}\right)^n \mathcal{H}^n(X_{t,s}(K)). \tag{4.10}$$

Moreover if we choose s such that $t-s$ is small enough

$$\begin{aligned}\mathcal{H}^n(X_{t,s}(K)) &\geq C_1 \mathcal{H}^n(K) - C_2(t-s)\int_K d\Delta_s u(t,\cdot) + O((t-s)^2) \\ &\geq -C_2(t-s)\int_K d\Delta_c u(t,\cdot) + O((t-s)^2) \\ &\geq C_2(t-s)\omega + O((t-s)^2) \\ &\geq \frac{C_2}{2}\omega^2,\end{aligned}$$

where we have used the fact that $\mathcal{H}^n(K) = 0$, that $\Delta_j u(t,K) \leq 0$, which is true due to semiconcavity, implies $\Delta_s u(t,K) \leq \Delta_c u(t,K)$, and $-\Delta_c u(t,K) \geq |D_c^2 u(t,\cdot)(K)| = \omega$. Thus

$$\mathcal{H}^n(X_{t,s}(K)) \geq \frac{C_2}{2}\omega^2. \tag{4.11}$$

Combining (4.10) with (4.11) we obtain

$$\mathcal{H}^n(K) \geq \left(\frac{1}{2}\right)^n \left(\frac{\delta}{t+\delta-s}\right)^n \frac{C_2}{2}\omega^2 > 0.$$

This is in contradiction with our hypothesis. □

We now have all the necessary Lemmas to prove the Theorem 1.1.

Proof. For $\varepsilon > 0$ sufficiently small such that Lemmas 4.4, 4.7, 4.8, and 4.10 hold, consider the functional F defined in (4.1) over the interval $[\tau, \tau+\varepsilon]$. F is bounded, and, from Lemma 4.4, F is a monotone function. Thus its points of discontinuity are at most countable.

We will prove that the presence of a Cantor part at a time t is related to a discontinuity of the functional F in t, hence there must be only a countable number of t's in $[\tau, \tau+\varepsilon]$ for which the Cantor part is negative.

Suppose there exists a t in $(\tau, \tau+\varepsilon)$ such that

$$|D_c^2 u(t, \Omega_t)| > 0,$$

then let A be the set of Lemma 4.10. Using Lemma 4.10-(iii), for any $\delta > 0$ we get

$$F(t+\delta) \le F(t) - \mathcal{H}^n(X_{t,\tau}(A)) \tag{4.12}$$

To compute $\mathcal{H}^n(X_{t,\tau}(A))$ call $\omega := |D_c^2 u(t,\cdot)|(A)$. As we saw in the previous lemma, if we choose $s \in [\tau, t)$ such that $t-s$ is small enough, we have

$$\mathcal{H}^n(X_{t,s}(A)) \ge \frac{C_2}{2}\omega^2.$$

Moreover for Lemma 4.8

$$\mathcal{H}^n(X_{t,\tau}(A)) \ge \left(\frac{1}{2}\right)^n \left(\frac{t-\tau}{t-s}\right)^n \mathcal{H}^n(X_{t,s}(A)).$$

Hence

$$\mathcal{H}^n(X_{t,\tau}(A)) \ge \left(\frac{1}{2}\right)^n \left(\frac{t-\tau}{t-s}\right)^n \frac{C_2}{2}\omega^2 \ge C\omega^2.$$

We can now use this estimate in (4.12) obtaining

$$F(t+\delta) \le F(t) - C\omega^2.$$

Letting $\delta \to 0$

$$\limsup_{\delta\to 0} F(t+\delta) < F(t).$$

Therefore t is a point of discontinuity for F, as we would like to prove.
□

REFERENCES

[1] G. Alberti and L. Ambrosio, *A geometrical approach to monotone functions in* $\mathbb{R}^n$, Math. Z., 2 (1999), pp. 259–316.

[2] L. Ambrosio and C. De Lellis, *A note on admissible solutions of 1D scalar conservation laws and 2D Hamilton-Jacobi equations*, J. Hyperbolic Differ. Equ., 31 (4) (2004), pp. 813–826.

[3] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford University Press, 2000.

[4] E.N. Barron, P. Cannarsa, R. Jensen, and C. Sinestrari, *Regularity of Hamilton-Jacobi equations when forward is backward*, Indiana Univ. Math. J., 48 (1999), pp. 385–409.

[5] S. Bianchini, C. De Lellis, and R. Robyr, *SBV regularity for Hamilton-Jacobi equations in* $\mathbb{R}^n$, Arch. Ration. Mech. Anal., 200 (2011), pp. 1003–1021.

[6] P. Cannarsa, A. Mennucci, and C. Sinestrari, *Regularity results for solutions of a class of Hamilton-Jacobi equations*, Arch. Ration. Mech. Anal., 140 (1997), pp. 197–223.

[7] P. Cannarsa and C. Sinestrari, *Viscosity solutions of Hamilton-Jacobi equations and optimal control problems*, Birkhäuser, Boston, 2004.

[8] P. Cannarsa and H.M. Soner, *On the singularities of the viscosity solutions to Hamilton-Jacobi equations*, Indiana Univ. Math. J., 36 (1987), pp. 501–524.

[9] M.G. Crandall, L.C. Evans, and P.-L. Lions, *Some properties of viscosity solutions of Hamilton-Jacobi equations*, Trans. Amer. Math. Soc., 282 (1984), pp. 487–502.

[10] M.G. Crandall and P.-L. Lions, *Viscosity solutions of Hamilton-Jacobi equations*, Trans. Amer. Math. Soc., 277 (1983), pp. 1–42.

[11] W.H. Fleming, *The Cauchy problem for a nonlinear first order partial differential equation*, J. Differential Equations, 5 (1969), pp. 515–530.

[12] W.H. Fleming and R.W. Rishel, *Deterministic and stochastic optimal control*, Springer, New York, 1975.

[13] W.H. Fleming and H.M. Soner, *Controlled Markov process and viscosity solutions*, Springer, Berlin, 1993.

[14] P.-L. Lions, *Generalized solutions of Hamilton-Jacobi equations*, Pitman, Boston, 1982.