Metric entropy for Hamilton-Jacobi equation with uniformly directionally convex Hamiltonian

Stefano Bianchini[(1)], Prerona Dutta[(2)] and Khai T. Nguyen[(3)]

[(1)] SISSA, via Beirut 2, IT-34014 Trieste, ITALY
[(2)] The Ohio State University, Columbus, OH 43210, USA
[(3)] North Carolina State University, Raleigh, NC 27695, USA

E-mails: bianchin@sissa.it, dutta.105@osu.edu, khai@math.ncsu.edu

Abstract

The present paper studies the BV-type regularity for viscosity solutions of the Hamilton-Jacobi equation

$$u_t(t,x) + H\big(D_x u(t,x)\big) = 0, \qquad (t,x) \in (0,\infty) \times \mathbb{R}^d,$$

with a coercive and uniformly directionally convex Hamiltonian H. More precisely, we establish a BV bound on the slope of backward characteristics $DH(D_x u(t,\cdot))$ starting at a positive time t. Relying on the BV bound, we quantify the metric entropy in $\mathbf{W}^{1,1}_{\mathrm{loc}}\left(\mathbb{R}^d\right)$ for the map S_t that associates to every given initial data $u_0 \in \mathbf{Lip}\big(\mathbb{R}^d\big)$, the corresponding solution $S_t u_0$. Finally, a counter example is constructed to show that both $D_x u(t,\cdot)$ and $DH(D_x u(t,\cdot))$ fail to be in BV_{loc} for a general strictly convex and coercive $H \in \mathcal{C}^2\left(\mathbb{R}^d\right)$.

Keywords: Hamilton-Jacobi equations, Hopf-Lax semigroup, Kolmogorov entropy, semiconcave functions, bounded total variation

MSC Subject classifications: 49L05, 47H20, 49L20, 47H0

1 Introduction

Consider a first-order Hamilton-Jacobi equation

$$u_t(t,x) + H\big(D_x u(t,x)\big) = 0, \qquad (t,x) \in (0,\infty) \times \mathbb{R}^d, \tag{1.1}$$

where $u : [0,+\infty)\times\mathbb{R}^d \to \mathbb{R}^d$, $D_x u = (u_{x_1},\dots,u_{x_d})$ and $H : \mathbb{R}^d \to \mathbb{R}$ is a Hamiltonian. Due to the nonlinear dependence of the characteristic speeds on the gradient of the solution, in general a classical solution u will develop singularities and the gradient $D_x u$ will become discontinuous in finite time. To cope with this difficulty, the concept of viscosity solution was introduced by Crandall and Lions in [11] to guarantee global existence, uniqueness and stability of the Cauchy problem, under suitable assumptions on the Hamiltonian H. In particular, assume that

(H1) $H \in \mathcal{C}^1(\mathbb{R}^d)$ is coercive and strictly convex, i.e., $\lim_{|p|\to\infty} \frac{H(p)}{|p|} = +\infty$ and

$$H(tp_1 + (1-t)p_2) \ < \ t \cdot H(p_1) + (1-t)H(p_2), \quad t \in (0,1),\ p_1, p_2 \in \mathbb{R}^d.$$

The Hamilton-Jacobi equation (1.1) generates a Hopf-Lax semigroup of viscosity solutions $\left\{S_t : \mathbf{Lip}\left(\mathbb{R}^d\right) \to \mathbf{Lip}\left(\mathbb{R}^d\right)\right\}_{t\geq 0}$ such that for every initial data $u_0 \in \mathbf{Lip}\left(\mathbb{R}^d\right)$, the corresponding unique viscosity solution of equation (1.1) with $u(0,x) = u_0(x)$ is computed by the Hopf-Lax representation formula

$$u(t,x) = S_t(u_0)(x) = \min_{y\in\mathbb{R}^d}\left\{u_0(y) + t\cdot L\left(\frac{x-y}{t}\right)\right\} \tag{1.2}$$

where L is the Legendre transform of H. In addition, if H is strongly convex, i.e., there exists a constant $\lambda > 0$ such that

$$D^2 H(p) \geqslant \lambda \cdot \mathbf{I}_d \quad \text{for all } p \in \mathbb{R}^d,$$

then $u(t,\cdot)$ is twice differentiable almost everywhere and $D_x u(t,\cdot)$ has locally bounded total variation. Thanks to Helly's compactness theorem, the map $S_t : \mathbf{Lip}(\mathbb{R}^d) \to \mathbf{Lip}(\mathbb{R}^d)$ is compact in $\mathbf{W}^{1,1}_{\text{loc}}\left(\mathbb{R}^d\right)$. A natural question arises on how to measure the degree of compactness of S_t. This involves using the ε-entropy, introduced by Kolmogorov and Tikhomirov in [14]:

Definition 1.1 *Let (E,ρ) be a metric space and F be a totally bounded subset of E. For $\varepsilon > 0$, let $\mathcal{N}_\varepsilon(F|E)$ be the minimal number of sets in a covering of F by subsets of E having diameter no larger than 2ε. Then the ε-entropy of F is defined as*

$$\mathcal{H}_\varepsilon(F|E) \ := \ \log_2 \mathcal{N}_\varepsilon(F|E).$$

In other words, it is the minimum number of bits needed to represent a point in a given set F in the space E, up to an accuracy ε with respect to the metric ρ. Such an approach stems from a conjecture of Lax in [15] for scalar conservation laws with uniformly convex fluxes. A complete answer to Lax's conjecture was provided in [5, 6, 12]. This study was also extended to scalar conservation laws with nonconvex fluxes in [8, 10] and to hyperbolic systems of conservation laws in [6, 7]. Recently, the first results on the ε-entropy for sets of viscosity solutions of (1.1) were obtained in [3, 4]. The authors proved that the minimal number of bits needed to represent a viscosity solution of (1.1) up to an accuracy ε with respect to the $\mathbf{W^{1,1}}$-distance is of the order ε^{-d} under the strongly convex condition on Hamiltonian H. There the main idea was to provide controllability results for Hamilton-Jacobi equations and a compactness result for a class of semiconcave functions. However, such a gain of BV regularity does not hold for (1.1) with a general strictly convex Hamiltonian H and the previous approach to finding ε-entropy of the solution set cannot be applied.

In this paper, we first study the fine regularity properties of viscosity solutions to (1.1) when H satisfies **(H1)** and the following assumption of uniformly directional convexity:

(H2) For every constant $r > 0$, it holds that

$$\inf_{p \neq q \in \overline{B}(0,r)} \left\langle \frac{DH(p) - DH(q)}{|DH(p) - DH(q)|}, \frac{p-q}{|p-q|} \right\rangle := \lambda_r > 0. \tag{1.3}$$

Notice that strong convexity on H implies **(H2)** but not vice versa (e.g., $H(p) = |p|^4$). Moreover, in the scalar case ($d = 1$), (1.3) holds for every $H \in \mathcal{C}^2(\mathbb{R})$ with $H'' > 0$, without requiring strong convexity. Furthermore, we refer to Remark 2.6 which gives a sufficient condition for **(H2)** in $\mathbb{R}^d$ with $d \geq 2$. By the Hopf-Lax representation formula (1.2), it is well known from [9] that the set of slopes of backward optimal rays through (t, x), denoted by

$$\mathbf{b}(t,x) = \left\{ \frac{x-y}{t} : y \in C_{t,x} \right\}, \quad C_{t,x} = \arg\min_{y \in \mathbb{R}^d} \left\{ u_0(y) + t \cdot L\left(\frac{x-y}{t}\right) \right\},$$

reduces to a singleton $\mathbf{b}(t,x) = \{DH(D_x u(t,x))\}$ for almost every $(t,x) \in [0,\infty) \times \mathbb{R}^d$ and can be viewed as an element in $\mathbf{L}^\infty(\mathbb{R}^d)$. Towards the sharp estimate on ε-entropy of the semigroup S_t, we establish a BV bound on $\mathbf{b}(t, \cdot)$.

Theorem 1.2 *Assume that H satisfies* **(H1)**-**(H2)**. *For every $t > 0$ and $u_0 \in \mathbf{Lip}(\mathbb{R}^d)$ with a Lipschitz constant $M > 0$, $\mathbf{b}(t, \cdot)$ has locally bounded total variation and its total variation $|D_x \mathbf{b}(t, \cdot)|$ over an open and bounded set $\Omega \subset \mathbb{R}^d$ of finite perimeter is bounded by*

$$|D_x \mathbf{b}(t,\cdot)|(\Omega) \ \leq \ \frac{1}{\gamma_M} \cdot \left(\Lambda_M + \frac{\operatorname{diam}(\Omega)}{t} \right) \cdot \mathcal{H}^{d-1}(\partial\Omega) + \frac{\sqrt{d}}{t} \cdot |\Omega|$$

for some constants $\gamma_M, \Lambda_M > 0$ depending on M and H.

Intuitively, the uniformly directional convexity of H yields a bound on the directional derivatives of $\mathbf{b}(t, \cdot)$ in terms of $\operatorname{div}_x(\mathbf{b}(t, \cdot))$. Indeed, to prove Theorem 1.2, we provide an upper bound on the quotient $|D_x \mathbf{b}(t, \cdot)|/|\operatorname{div}_x \mathbf{b}(t, \cdot)|$ for a suitable sequence of approximate solutions, which converges uniformly and monotonically to the given solution. In turn, the approximations of $\mathbf{b}$ will converge (in the sense that their graphs converge with respect to the Hausdorff distance). As a consequence of Theorem 1.2, for every $t > 0$, the map $S_t : \mathbf{Lip}(\mathbb{R}^d) \to \mathbf{Lip}(\mathbb{R}^d)$ is compact in $\mathbf{W}^{1,1}_{\text{loc}}(\mathbb{R}^d)$.

In the second part of the paper, we shall use the bound on the total variation of $\mathbf{b}(t, \cdot)$ to quantify the compactness of S_t for $t > 0$. More precisely, given constants $m, M, R, T > 0$, consider the set of initial data

$$\mathcal{U}_{[m,M]} := \left\{ \bar{u} \in \mathbf{Lip}(\mathbb{R}^d) : |\bar{u}(0)| \leq m, \operatorname{Lip}[\bar{u}] \leq M \right\}.$$

We establish upper and lower estimates for the ε-entropy of the following solution set at time T

$$S_{T,R}\left(\mathcal{U}_{[m,M]}\right) := \left\{ v_{\llcorner \Box_R} \ : \ v \in S_T\left(\mathcal{U}_{[m,M]}\right) \right\}$$

with $\Box_R = (-R, R)^d$ and $v_{\llcorner \Box_R}$ denoting the restriction of v on $\Box_R$.

Theorem 1.3 *Assume that $H \in \mathcal{C}^2(\mathbb{R}^d)$ and satisfies* **(H1)**-**(H2)**. *There exist constants $C_1, C_2, R_1, R_2 > 0$ such that for every $\varepsilon > 0$ sufficiently small,*

$$C_1 \cdot \left(\Phi_M\left(\frac{\varepsilon}{R_1}\right)\right)^{-d} \leq \mathcal{H}_\varepsilon\left(S_{T,R}(\mathcal{U}_{[m,M]})\Big|\mathbf{W}^{1,1}(\square_R)\right) \leq C_2 \cdot \left(\Psi_M\left(\frac{\varepsilon}{R_2}\right)\right)^{-d}.$$

Here, Φ_M, Ψ_M are strictly increasing functions which depend on H and will be explicitly defined in Section 4. In Remark 4.2 we particularly consider $H(p) = |p|^{2\alpha}$ for $\alpha \geq 1$, which satisfies **(H1)**-**(H2)** and show that in this case

$$\mathcal{H}_\varepsilon\left(S_{T,R}(\mathcal{U}_{[m,M]})\big|\mathbf{W}^{1,1}(\square_R)\right) \approx \varepsilon^{-(2\alpha-1)d}.$$

Using Theorem 1.2 and a result in [13], the ε-entropy of the sets of slopes of optimal rays starting at time T in $\mathbf{L}^1_{\text{loc}}(\mathbb{R}^d)$ is found to be of the order ε^{-d}. Thus, to achieve the upper bound in the above theorem, we establish a quantitative relation (depending on the nonlinearity of H) between the $\mathbf{W}^{1,1}$-distance of two solutions and the $\mathbf{L}^1$-distance of slopes of two corresponding optimal rays. Finally, towards the derivation of the lower bound on $\mathcal{H}_\varepsilon\left(S_{T,R}(\mathcal{U}_{[m,M]})\Big|\mathbf{W}^{1,1}(\square_R)\right)$, we study a controllability result for (1.1). In particular, we show that a solution to (1.1) with a semiconvex initial condition preserves the semiconvexity on a given time interval, provided the semiconvexity constant of the initial data is sufficiently small in absolute value.

The remainder of this paper is organized as follows. In Section 2, we collect preliminary results and definitions related to semiconcave functions, BV functions and Hamilton-Jacobi equations. In Section 3, we prove the BV-type regularity for viscosity solutions. Relying on this result, in Section 4 we establish a sharp estimate on the ε-entropy of the map S_T. Finally in Section 5, we construct a counter-example to show that if $H \in \mathcal{C}^2(\mathbb{R}^d)$ satisfies **(H1)** but not **(H2)** then both $D_x u(t,\cdot)$ and $\mathbf{b}(t,\cdot)$ fail to be in BV_{loc} in general.

2 Notation and preliminaries

Given a positive integer d and a measurable set $\Omega \subseteq \mathbb{R}^d$, throughout the paper we shall denote by

- $|\cdot|$, the Euclidean norm in $\mathbb{R}^d$ and

$$B_d(x,R) = \{y \in \mathbb{R}^d : |x-y| < R\} \qquad \text{for all } R > 0;$$

- $\langle\cdot,\cdot\rangle$, the Euclidean inner product in $\mathbb{R}^d$;
- $\otimes$, the tensor product;
- $\partial\Omega$, the boundary of Ω;
- $[x,y]$, the segment joining two points $x, y \in \mathbb{R}^d$;
- $\#S$, the number of elements in any finite set S;
- $\text{Vol}(D)$, the Lebesgue measure of a measurable set $D \subset \mathbb{R}^d$;

- $\omega_d := \text{Vol}(B_d(0,1))$, the Lebesgue measure of the unit ball in $\mathbb{R}^d$;
- $\mathbf{L}^1(\Omega)$, the Lebesgue space of all (equivalence classes of) summable real-valued functions on Ω, equipped with the usual norm $\|\cdot\|_{\mathbf{L}^1(\Omega)}$ (we shall use the same symbol in case u is vector-valued);
- $\mathbf{L}^\infty(\Omega)$, the space of all essentially bounded real-valued functions on Ω and $\|u\|_{\mathbf{L}^\infty(\Omega)}$ is the essential supremum of a function $u \in \mathbf{L}^\infty(\Omega)$ (we shall use the same symbol in case u is vector-valued);
- $\mathbf{W}^{1,1}(\Omega)$, the Sobolev space of functions with summable first order distributional derivatives and $\|\cdot\|_{\mathbf{W}^{1,1}(\Omega)}$ is its norm;
- $\mathcal{C}^1(\Omega)$, the space of continuously differentiable real valued functions on Ω;
- $\mathcal{C}^1_c(\Omega, \mathbb{R}^d)$, the space of continuously differentiable functions $u : \Omega \to \mathbb{R}^d$ with a compact support;
- $\mathbf{Lip}(\Omega)$, the space of all Lipschitz functions $f : \Omega \to \mathbb{R}$ and $\text{Lip}[f]$ is the Lipschitz seminorm of f;
- $\mathcal{H}^k(E)$, the k-dimensional Hausdorff measure of $E \subset \mathbb{R}^d$;
- For any function f, the function $f_{\llcorner\Omega}$ is the restriction of f on Ω;
- $\mathbf{I}_d$, the identity matrix of size d;
- $\lfloor a \rfloor := \max\{z \in \mathbb{Z} : z \leq a\}$, the integer part a.

2.1 Semiconcave and BV functions in $\mathbb{R}^d$

2.1.1 Semiconcave functions

Let us recall some basic definitions and properties of semiconcave (semiconvex) functions in $\mathbb{R}^d$. We refer to [9] for a general introduction to the respective theories.

Definition 2.1 *A continuous function $u : \Omega \to \mathbb{R}$ is semiconcave with a semiconcavity constant K if for all $x, h \in \mathbb{R}^d$ with $[x-h, x+h] \subset \Omega$, it holds that*

$$u(x+h) + u(x-h) - 2u(x) \leq K \cdot |h|^2.$$

We say that

- *u is semiconvex (with constant $-K$) if $-u$ is semiconcave (with constant K);*
- *u is locally semiconcave (semiconvex) if u is semiconcave (semiconvex) in every compact set $A \subset \Omega$.*

We denote the distributional gradient of a semiconcave function u by Du and for every $x \in \Omega$ with $\Omega \subseteq \mathbb{R}^d$ open, we define the superdifferential and the subdifferential of u at x respectively by

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

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

It is clear that $D^{\pm}u(x)$ is convex and $D^-u(x) = -D^+(-u)(x)$ for all $x \in \Omega$. From [9, Proposition 3.3.4, Proposition 3.3.10], the superdifferential of a semiconcave function enjoys the following properties.

Proposition 2.1 *Given a convex and open set $\Omega \subseteq \mathbb{R}^d$, let $u : \Omega \to \mathbb{R}$ be semiconcave with a semiconcavity constant K. Then*

(i) The superdifferential $D^+u(x)$ is a compact, convex, nonempty set for all $x \in \Omega$. Moreover, the set-valued map $x \mapsto D^+u(x)$ is upper semicontinuous;

(ii) $D^+u(x)$ is a singleton if and only if u is differentiable at x. Furthermore, if $D^+u(x)$ is a singleton for all $x \in \Omega$, then $u \in \mathcal{C}^1(\Omega)$;

(iii) For every $x_1, x_2 \in \Omega$, it holds that

$$\langle p_2 - p_1, x_2 - x_1 \rangle \leq K \cdot |x_2 - x_1|^2, \qquad p_i \in D^+u(x_i), i \in \{1, 2\}.$$

From (ii) if u is both locally semiconcave and locally semiconvex then u is in $\mathcal{C}^1(\Omega)$ as shown in [9, Corollary 3.3.8]. This is crucial to prove further regularity for viscosity solutions in Proposition 2.2 which allows us to construct a backward smooth solution of (1.1). To complete this part, for every constant $r, K > 0$, let us define the set

$$\mathcal{SC}_{[r,K]} := \left\{ v \in \mathbf{Lip}(\mathbb{R}^d) : \mathrm{Lip}[v] \leq r \text{ and } v \text{ is semiconcave with constant } K \right\}. \tag{2.1}$$

From the proof of [3, Proposition 10], one can easily obtain a lower bound on the ε-entropy for the set $\left\{ Dv_{\llcorner \Box_R} : v \in \mathcal{SC}_{[r,K]} \right\}$ in $\mathbf{L}^1(\Box_R)$ which will be used to establish a lower estimate on the ε-entropy of a set of viscosity solutions in subsection 4.2.

Corollary 2.2 *Given any $r, R, K > 0$, for every $\varepsilon > 0$ sufficiently small, there exists a subset $\mathcal{G}^R_{[r,K]}$ of $\mathcal{SC}_{[r,K]}$ such that*

$$\#\mathcal{G}^R_{[r,K]} \;\geq\; 2^{\beta_{[R,K]} \cdot \varepsilon^{-d}}, \qquad \beta_{[R,K]} \;=\; \frac{1}{3^d 2^{d^2+4d+3} \ln 2} \cdot \left(\frac{K \omega_d R^{d+1}}{(d+1)} \right)^d$$

and

$$\left\| Dv_{\llcorner \Box_R} - Dw_{\llcorner \Box_R} \right\|_{\mathbf{L}^1(\Box_R)} \;\geq\; 2\varepsilon \qquad \textit{for all } v \neq w \in \mathcal{G}^R_{[r,K]}.$$

2.1.2 Functions of bounded total variation

Let us now introduce the concept of functions of bounded variation. We refer to [2] for a comprehensive analysis on this topic.

Definition 2.3 *The function* $u \in \mathbf{L}^1(\Omega)$ *is a function of bounded variation on* $\Omega \subseteq \mathbb{R}^d$ *and said to be in* $BV(\Omega, \mathbb{R}^m)$*, if the distributional derivative of* u*, denoted by* Du*, is an* $m \times d$ *matrix of finite measures* $D_i u^\alpha$ *in* Ω *satisfying*

$$\sum_{\alpha=1}^{m} \int_\Omega u^\alpha \, div\, \varphi^\alpha \; dx \;=\; -\sum_{\alpha=1}^{m} \sum_{i=1}^{d} \int_\Omega \varphi_i^\alpha dD_i u^\alpha \quad \text{for all } \varphi \in \left[\mathcal{C}_c^1(\Omega, \mathbb{R}^d)\right]^m.$$

We denote by $|Du|(\Omega)$ *the total variation of* u *over* Ω*, i.e.,*

$$|Du|(\Omega) \;=\; \sup\left\{ \sum_{\alpha=1}^{m} \int_\Omega u^\alpha \, div\, \varphi^\alpha \; dx : \varphi \in \left[\mathcal{C}_c^1(\Omega, \mathbb{R}^d)\right]^m, \|\varphi\|_\infty \leq 1 \right\}.$$

We recall a Poincaré-type inequality for bounded total variation functions on convex domains that will be used in the paper. This result is based on [1, Theorem 3.2].

Theorem 2.4 *(Poincaré inequality) Let* $\Omega \subset \mathbb{R}^d$ *be an open, bounded, convex set with Lipschitz boundary. For any* $u \in BV(\Omega, \mathbb{R})$*, it holds that*

$$\int_\Omega \left| u(x) - u_\Omega \right| \; dx \;\leq\; \frac{\mathrm{diam}(\Omega)}{2} \cdot |Du|(\Omega)$$

where

$$u_\Omega \;=\; \frac{1}{\mathrm{Vol}(\Omega)} \cdot \int_\Omega u(x) \; dx$$

is the mean value of u *over* Ω*.*

To complete this subsection, let us recall a result on the metric entropy for a class of functions with bounded total variation which will be used in subsection 4.1. For every $R, M, V > 0$, we consider a class of uniformly bounded total variation functions on $\Box_R$

$$\mathcal{F}_{[R,M,V]} \;=\; \left\{ f : \Box_R \to \mathbb{R}^d : \|f\|_{\mathbf{L}^\infty(\Box_R)} \leq M, |Df|(\Box_R) \leq V \right\}. \tag{2.2}$$

By a slight modification in the proof of [13, Theorem 1], one can obtain the following upper bound on the ε-entropy of $\mathcal{F}_{[R,M,V]}$ in $\mathbf{L}^1(\Box_R)$.

Corollary 2.5 *For every* $0 < \varepsilon < \min\left\{ \frac{6RV^2}{3V+2M}, 2V\left(\frac{RV}{M}\right)^{\frac{1}{d}} \right\}$*, it holds that*

$$\mathcal{H}_\varepsilon\left(\mathcal{F}_{[R,M,V]} \;\Big|\; \mathbf{L}^1(\Box_R) \right) \;\leq\; 48\sqrt{d} \cdot \left(\frac{6d\sqrt{d}RV}{\varepsilon} \right)^d. \tag{2.3}$$

Proof. By the definition of ε-entropy, we have

$$\mathcal{H}_\varepsilon\left(\mathcal{F}_{[R,M,V]} \;\Big|\; \mathbf{L}^1(\Box_R) \right) \;\leq\; d \cdot \mathcal{H}_{\frac{\varepsilon}{d}}\left(\mathcal{F}^1_{[R,M,V]} \;\Big|\; \mathbf{L}^1(\Box_R) \right) \tag{2.4}$$

with

$$\mathcal{F}^1_{[R,M,V]} \;=\; \left\{ f : \Box_R \to \mathbb{R} : \|f\|_{\mathbf{L}^\infty(\Box_R)} \leq M, |Df|(\Box_R) \leq V \right\}.$$

Consider a class of real-valued bounded total variation functions

$$\mathcal{B}_{[R,M,V]} \;=\; \{g:[0,R]\to[0,M] : |Dg|([0,R])\le V\}\,.$$

From [13, Lemma 2.3], for every $0<\varepsilon<\dfrac{RV}{3}$, one has

$$\mathcal{N}_\varepsilon\left(\mathcal{B}_{[R,\frac{9}{8}V,V]}\Big|\mathbf{L}^1([0,R])\right) \;\le\; 2^{\frac{17RV}{\varepsilon}}$$

and this implies that

$$\mathcal{N}_\varepsilon\left(\mathcal{B}_{[R,M,V]}\Big|\mathbf{L}^1([0,R])\right) \;\le\; \frac{8M}{V}\cdot\mathcal{N}_\varepsilon\left(\mathcal{B}_{[R,\frac{9}{8}V,V]}\Big|\mathbf{L}^1([0,R])\right) \;\le\; \frac{8M}{V}\cdot 2^{\frac{17RV}{\varepsilon}}.$$

In particular, for every $0<\varepsilon<\dfrac{RV^2}{3V+M}$ such that $\dfrac{8M}{V}\le 2^{\frac{RV}{\varepsilon}}$, it holds that

$$\mathcal{H}_\varepsilon\left(\mathcal{B}_{[R,M,V]}\Big|\mathbf{L}^1([0,R])\right) \;=\; \log_2\left(\mathcal{N}_\varepsilon\left(\mathcal{B}_{[R,M,V]}\Big|\mathbf{L}^1([0,R])\right)\right) \;\le\; \frac{18RV}{\varepsilon}.$$

Using the above estimate, one can follow the same argument as in the proof of [13, Theorem 3.1] to obtain that for every $0<\varepsilon<\min\left\{\dfrac{6RV^2}{3V+2M}, 2V\left(\dfrac{RV}{M}\right)^{\frac{1}{d}}\right\}$, it holds that

$$\mathcal{H}_\varepsilon\left(\mathcal{F}^1_{[R,M,V]}\;\Big|\;\mathbf{L}^1(\Box_R)\right) \;\le\; \frac{48}{\sqrt{d}}\cdot\left(\frac{6\sqrt{d}RV}{\varepsilon}\right)^d$$

and then (2.4) yields (2.3). □

2.2 Semigroup of Hamilton-Jacobi equation

Consider the Hamilton-Jacobi equation (1.1) under the assumptions **(H1)-(H2)**. Without loss of generality, we shall assume that the Hamiltonian satisfies further conditions

$$H(0)=0 \qquad\text{and}\qquad DH(0)=0, \tag{2.5}$$

otherwise the transformations $x\mapsto x+tDH(0)$, $u(t,\cdot)\mapsto u(t,x)+t\cdot H(0)$ and $H(p)\mapsto H(p)-\langle DH(0),p\rangle$ reduce the general case to this one. Before recalling the concept of viscosity solution to (1.1), let us give a sufficient condition for the assumption **(H2)**.

Remark 2.6 *Let H be in $\mathcal{C}^2(\mathbb{R}^d)$. Assume that there exists a constant $\lambda>0$ such that*

$$D^2H(p) \;=\; |D^2H(p)|\cdot A(p), \qquad A(p)\;\ge\;\lambda\cdot\mathbf{I}_d, \tag{2.6}$$

with $A(p)$ being a $d\times d$ matrix and $|D^2H(p)|$ denoting the matrix norm of $D^2H(p)$. Then H satisfies **(H2)**.

Proof. For any $p \neq q \in \mathbb{R}^d$, by mean value theorem, it holds that

$$\begin{aligned} DH(p) - DH(q) = & \int_0^1 D^2H(tp + (1-t)q) \cdot (p-q)dt \\ = & \left[\int_0^1 A(tp+(1-t)q)\big|D^2H(tp+(1-t)q)\big|dt\right] \cdot (p-q) \end{aligned}$$

and

$$|DH(p) - DH(q)| \leq |p-q| \cdot \int_0^1 |D^2H(tp+(1-t)q)|dt.$$

Using (2.6), we estimate

$$\begin{aligned} \langle DH(p) - DH(q), p-q\rangle = & \int_0^1 \Big[(p-q)^T A(tp+(1-t)q)(p-q)\Big] \cdot \big|D^2H(tp+(1-t)q)\big|dt \\ \geq & \ \lambda \cdot |p-q|^2 \int_0^1 \big|D^2H(tp+(1-t)q)\big|dt \\ \geq & \ \lambda \cdot |DH(p) - DH(q)| \cdot |p-q| \end{aligned}$$

and this implies (1.3). □

As we mentioned in the introduction, classical smooth solutions of (1.1) in general break down and Lipschitz continuous functions that satisfy (1.1) almost everywhere together with a given initial condition are not unique. To handle this problem, the following concept of a generalized solution was introduced in [11] to guarantee global existence and uniqueness results.

Definition 2.7 *(Viscosity solution) We say that a continuous function* $u : [0,T]\times\mathbb{R}^d \to \mathbb{R}^d$ *is a viscosity solution of* (1.1) *if:*

(1) *u is a viscosity subsolution of (1.1), i.e., for every point $(t_0, x_0) \in (0,T) \times \mathbb{R}^d$ and test function $v \in \mathcal{C}^1\big((0,+\infty)\times\mathbb{R}^d\big)$ such that $u - v$ has a local maximum at (t_0, x_0), it holds that*

$$v_t(t_0,x_0) + H\big(D_x v(t_0,x_0)\big) \leqslant 0\,,$$

(2) *u is a viscosity supersolution of (1.1), i.e., for every point $(t_0, x_0) \in (0,T) \times \mathbb{R}^d$ and test function $v \in \mathcal{C}^1\big((0,+\infty)\times\mathbb{R}^d\big)$ such that $u - v$ has a local minimum at (t_0, x_0), it holds that*

$$v_t(t_0,x_0) + H\big(D_x v(t_0,x_0)\big) \geqslant 0\,.$$

By the alternative equivalent definition of viscosity solution expressed in terms of the subdifferential and superdifferential of the function as in [11] and because of Proposition 2.1 one immediately sees that every $\mathcal{C}^1$ solution of (1.1) is also a viscosity solution of (1.1). On the other hand, if u is a viscosity solution of (1.1) then u satisfies the equation at every point of differentiability. Let us state a result on further regularity for viscosity solutions proved in [3, Proposition 3] which says that smoothness in the pair (t, x) follows from smoothness in the second variable.

Proposition 2.2 *Let u be a viscosity solution of (1.1) in $[0,T]\times\mathbb{R}^d$. If $u(t,\cdot)$ is both locally semiconcave and semiconvex in $\mathbb{R}^d$ for all $t\in(0,T)$ then u is a $\mathcal{C}^1$ solution of (1.1) in $(0,T]\times\mathbb{R}$.*

The viscosity solution of the Hamilton-Jacobi equation (1.1) with initial data $u(0,\cdot)=u_0\in\mathbf{Lip}(\mathbb{R}^d)$ can be represented as the value function of a classical problem in calculus of variations, which admits the Hopf-Lax representation formula

$$u(t,x)=\min_{y\in\mathbb{R}^d}\left\{t\cdot L\Big(\frac{x-y}{t}\Big)+u_0(y)\right\},\qquad t>0, x\in\mathbb{R}^d, \tag{2.7}$$

where $L\in\mathcal{C}^1\left(\mathbb{R}^d\right)$ denotes the Legendre transform of H, defined by

$$L(q):=\max_{p\in\mathbb{R}^d}\big\{p\cdot q-H(p)\big\},\qquad q\in\mathbb{R}^d. \tag{2.8}$$

The main properties of viscosity solutions defined by the Hopf-Lax formula, which are of interest to this paper are recalled below (cfr. [9, Section 1.1, Section 6.4]).

Proposition 2.3 *Let u be the viscosity solution of (1.1) on $[0,+\infty)\times\mathbb{R}^d$, with continuous initial data u_0, defined by (2.7). Then the following hold true:*

(i) ***Functional identity****: For all $x\in\mathbb{R}^d$ and $0\leqslant s<t$, it holds that*

$$u(t,x)=\min_{y\in\mathbb{R}^d}\left\{u(s,y)+(t-s)\cdot L\Big(\frac{x-y}{t-s}\Big)\right\}.$$

(ii) ***Differentiability of u and uniqueness****: (2.7) admits a unique minimizer y_x if and only if $u(t,\cdot)$ is differentiable at x. In this case we have*

$$y_x=x-t\cdot DH\big(D_xu(t,x)\big),\qquad D_xu(t,x)\in D^-u_0(y_x).$$

(iii) ***Dynamic programming principle****: Let $t>s>0$, $x\in\mathbb{R}^d$, assume that y is a minimizer for (2.7) and define $z=\frac{s}{t}x+\left(1-\frac{s}{t}\right)y$. Then y is the unique minimizer over $\mathbb{R}^d$ of*

$$w\mapsto s\cdot L\Big(\frac{z-w}{s}\Big)+u_0(w),\qquad w\in\mathbb{R}^d.$$

As a consequence, the family of nonlinear operators $\big\{S_t:\mathbf{Lip}(\mathbb{R}^d)\to\mathbf{Lip}(\mathbb{R}^d)\big\}_{t\geq0}$ defined by the Hopf-Lax representation formula, i.e., $S_0u_0=u_0$ and

$$S_tu_0(x)=\min_{y\in\mathbb{R}^d}\left\{t\cdot L\Big(\frac{x-y}{t}\Big)+u_0(y)\right\},\qquad t>0\,,\ x\in\mathbb{R}^d, \tag{2.9}$$

enjoys the following properties:

(i) For every $u_0\in\mathbf{Lip}(\mathbb{R}^d)$, $u(t,x):=S_tu_0(x)$ provides the unique viscosity solution of the Cauchy problem (1.1) with initial data $u(0,\cdot)=u_0$.

(ii) (Semigroup property)

$$S_{t+s}u_0=S_t\,S_su_0,\qquad t,s\geq0, u_0\in\mathbf{Lip}(\mathbb{R}^d).$$

(iii) (Translation) For every constant $c\in\mathbb{R}$ we have that

$$S_t(u_0+c)=S_tu_0+c,\qquad t\geqslant0, u_0\in\mathbf{Lip}(\mathbb{R}^d). \tag{2.10}$$

3 BV bound on $\mathbf{b}(t,\cdot)$

Throughout this section, we shall assume that the Hamiltonian H satisfies **(H1)-(H2)** and (2.5). For a given initial datum $u_0 \in \mathbf{Lip}(\mathbb{R}^d)$ with $\text{Lip}[u_0] \leq M$, let u be the solution of (1.1) with $u(0,\cdot) = u_0$ and

$$\mathbf{b}(t,x) = \left\{ \frac{x-y}{t} : y \in C_{t,x} \right\}, \qquad C_{t,x} = \arg\min_{y\in\mathbb{R}^d} \left\{ u_0(y) + t \cdot L\left(\frac{x-y}{t}\right) \right\}. \tag{3.1}$$

It is well known from [9] that $\mathbf{b}(t,x) \subseteq DH(D^+u(t,x))$ and

$$\|\mathbf{b}(t,\cdot)\|_{\mathbf{L}^\infty(\mathbb{R}^d)} \leq \Lambda_M \ := \ \max\{|q| : L(q) \leq M|q|\}. \tag{3.2}$$

Indeed, to achieve (3.2) one observes for any $y_x \in C_{t,x}$ that

$$L\left(\frac{x-y_x}{t}\right) \ \leq \ \frac{u_0(x) - u_0(y_x)}{t} \ \leq \ M \cdot \frac{|x-y_x|}{t}.$$

In order to establish a BV bound on $\mathbf{b}(t,\cdot)$, we approximate u by a monotone decreasing sequence of continuous functions $u_n : [0,+\infty) \times \mathbb{R}^d \to \mathbb{R}$ defined by

$$u_n(t,x) := \min_{y\in\mathcal{Z}_n} \left\{ u_0(y) + t \cdot L\left(\frac{x-y}{t}\right) \right\}, \qquad \mathcal{Z}_n := 2^{-n}\mathbb{Z}^d. \tag{3.3}$$

Considering the associated set of slopes of backward optimal rays through (t,x)

$$\mathbf{b}_n(t,x) = \left\{ \frac{x-y}{t} : y \in C^n_{t,x} \right\}, \qquad C^n_{t,x} = \arg\min_{y\in\mathcal{Z}_n} \left\{ u_0(y) + t \cdot L\left(\frac{x-y}{t}\right) \right\}, \tag{3.4}$$

we prove the following lemma.

Lemma 3.1 *For every $t > 0$, it holds that*

$$\lim_{n\to\infty} \|u_n(t,\cdot) - u(t,\cdot)\|_\infty = 0 \qquad \text{and} \qquad \limsup_{n\to\infty} \|\mathbf{b}_n(t,\cdot)\|_{\mathbf{L}^\infty(\mathbb{R}^d)} \ \leq \ \Lambda_M. \tag{3.5}$$

Proof. 1. Fix $n \geq 1$ and $(t,x) \in (0,\infty) \times \mathbb{R}^d$. Pick any $\bar{y} \in C^n_{t,x}$, let $\bar{z}$ be in $\mathcal{Z}_n$ such that $|\bar{z} - \bar{y}| \leq \sqrt{d}2^{-n+1}$. From (3.2) and (3.3), we estimate

$$\begin{aligned} |u(t,x) - u_n(t,x)| \ &\leq \ |u_0(\bar{z}) - u_0(\bar{y})| + t \cdot \left| L\left(\frac{x-\bar{z}}{t}\right) - L\left(\frac{x-\bar{y}}{t}\right) \right| \\ &\leq \ \left(M + \max_{|q| \leq \Lambda_M + \frac{\sqrt{d}}{2^{n-1}\cdot t}} |DL(q)| \right) \cdot |\bar{z} - \bar{y}| \end{aligned}$$

and this yields the first equality in (3.5).

2. For every $y_{x,n} \in C^n_{t,x}$ and $x' \in \mathbb{R}^d$, it holds that

$$\begin{aligned} u_n(t,x') - u_n(t,x) \ &\leq \ t \cdot \left[L\left(\frac{x'-y_{x,n}}{t}\right) - L\left(\frac{x-y_{x,n}}{t}\right) \right] \\ &\leq \ DL\left(\frac{x-y_{x,n}}{t}\right) \cdot (x'-x) + O(|x'-x|). \end{aligned}$$

From (3.3)-(3.4), there exists $\bar{x} \in \mathcal{Z}_n$ with $|x - \bar{x}| \leq \sqrt{d}2^{-n+1}$ such that

$$u_0(y_{x,n}) + t \cdot L\left(\frac{x - y_{x,n}}{t}\right) \leq u_0(\bar{x}) + t \cdot L\left(\frac{x - \bar{x}}{t}\right).$$

Recalling that $\mathrm{Lip}[u_0] \leq M$, we get

$$L\left(\frac{x - y_{x,n}}{t}\right) \leq M \cdot \left|\frac{\bar{x} - y_{x,n}}{t}\right| + \max_{|q| \leq \frac{\sqrt{d}}{2^{n-1} \cdot t}} L(q).$$

From (2.5), it holds that $L(0) = 0$ and thus the above estimate yields the second part of (3.5). □

Now we restate and prove Theorem 1.2, which is our first main theorem.

Theorem 3.2 *Assume that H satisfies* **(H1)-(H2)** *and (2.5). For every $t > 0$ and $u_0 \in \mathbf{Lip}(\mathbb{R}^d)$ with $\mathrm{Lip}[u_0] \leq M$ for some $M \geq 0$, the function $\mathbf{b}(t, \cdot)$ has locally bounded total variation and its total variation $|D_x\mathbf{b}(t, \cdot)|$ over an open and bounded set $\Omega \subset \mathbb{R}^d$ of finite perimeter is bounded by*

$$|D_x\mathbf{b}(t, \cdot)|(\Omega) \;\leq\; \frac{1}{\gamma_M} \cdot \left(\Lambda_M + \frac{\mathrm{diam}(\Omega)}{t}\right) \cdot \mathcal{H}^{d-1}(\partial\Omega) + \frac{\sqrt{d}}{t} \cdot |\Omega| \tag{3.6}$$

with $\gamma_M := \inf_{r > \max_{|q| \leq \Lambda_M} |DL(q)|} \lambda_r$ and λ_r being as in (1.3).

Proof. The proof is divided into three steps.

Step 1. Consider the sequence of approximate solutions defined in (3.3). Fixing $n \geq 1$ and $t > 0$, we write $\mathcal{Z}_n = \{y_1, y_2, \ldots, y_k, \ldots\}$. For any $i \neq j$, the set

$$\mathcal{O}_{i,j} = \left\{x \in \mathbb{R}^d : u_0(y_i) + t \cdot L\left(\frac{x - y_i}{t}\right) < u_0(y_j) + t \cdot L\left(\frac{x - y_j}{t}\right)\right\}$$

is an open subset of $\mathbb{R}^d$ with a $\mathcal{C}^1$-boundary

$$\Gamma_{i,j} = \left\{x \in \mathbb{R}^d : u_0(y_i) + t \cdot L\left(\frac{x - y_i}{t}\right) = u_0(y_j) + t \cdot L\left(\frac{x - y_j}{t}\right)\right\}.$$

Set $\mathcal{V}_i := \bigcup_{j \neq i, j \geq 1} \mathcal{O}_{i,j}$. From (3.3) and (3.4), we have

$$\mathbf{b}_n(t, x) = \frac{x - y_i}{t}, \qquad x \in \mathcal{V}_i,\ i \geq 1. \tag{3.7}$$

In particular, $\mathbf{b}_n(t, \cdot)$ is in $BV_{loc}(\mathbb{R}^d)$ and

$$\begin{cases} D_x\mathbf{b}_n(t, x) &= \dfrac{\mathbf{I}_d}{t} \cdot \mathcal{H}^d + \dfrac{1}{2t} \cdot \displaystyle\sum_{i \neq j} (y_j - y_i) \otimes \nu_i(x) \mathcal{H}^{d-1}_{\llcorner \partial\mathcal{V}_i \cap \partial\mathcal{V}_j}, \\ \mathrm{div}_x\mathbf{b}_n(t, x) &= \dfrac{d}{t} \cdot \mathcal{H}^d + \dfrac{1}{2t} \cdot \displaystyle\sum_{i \neq j} \langle y_j - y_i, \nu_i(x)\rangle\, \mathcal{H}^{d-1}_{\llcorner \partial\mathcal{V}_i \cap \partial\mathcal{V}_j}, \end{cases} \tag{3.8}$$

where $\nu_i(x)$ is the inner normal vector to $\mathcal{V}_i$ and is computed by

$$\nu_i(x) \;=\; \frac{DL\left(\frac{x-y_j}{t}\right) - DL\left(\frac{x-y_i}{t}\right)}{\left|DL\left(\frac{x-y_j}{t}\right) - DL\left(\frac{x-y_i}{t}\right)\right|} \qquad \text{for } \mathcal{H}^{d-1} \textit{ a.e. } x \in \partial\mathcal{V}_i \bigcap \partial\mathcal{V}_j.$$

Given an open and bounded set $\Omega \subset \mathbb{R}^d$ of finite perimeter, one gets from (3.8) that

$$\left|D_x \mathbf{b}_n(t,\cdot) - \frac{\mathbf{I}_d}{t}\right|(\Omega) \;\leq\; \frac{1}{2t}\cdot \sum_{i\neq j} |y_j - y_i| \cdot \mathcal{H}^{d-1}(\Omega \bigcap \partial\mathcal{V}_i \bigcap \partial\mathcal{V}_j). \tag{3.9}$$

For a fixed $x \in \Omega\bigcap\partial\mathcal{V}_i\bigcap\partial\mathcal{V}_j$, setting $p_i := DL\left(\frac{x-y_i}{t}\right)$ and $p_j := DL\left(\frac{x-y_j}{t}\right)$, we have

$$\nu_i(x) = \frac{p_j - p_i}{|p_j - p_i|}, \qquad y_j - y_i = (DH(p_i) - DH(p_j))t.$$

Recalling (1.3), (3.5), and (3.7), we obtain that

$$|y_i - y_j| \;\leq\; -\frac{1}{\lambda_{\beta_n}}\cdot \langle y_j - y_i, \nu_i(x)\rangle \tag{3.10}$$

with $\beta_n := \max_{|q|\leq \|\mathbf{b}_n(t,\cdot)\|_{\mathbf{L}^\infty(\mathbb{R}^d)}} |DL(q)|$ satisfying $\lim_{n\to\infty} \beta_n = \max_{|q|\leq\Lambda_M} |DL(q)|$. Thus, (3.8)-(3.9) yield

$$\left|D_x \mathbf{b}_n(t,\cdot) - \frac{\mathbf{I}_d}{t}\right|(\Omega) \;\leq\; \frac{1}{\lambda_{\beta_n}}\cdot \left|\mathrm{div}_x \mathbf{b}_n(t,\cdot) - \frac{d}{t}\right|(\Omega). \tag{3.11}$$

Step 2. Let us now provide a bound on $\left|\mathrm{div}_x \mathbf{b}_n(t,\cdot) - \frac{d}{t}\right|(\Omega)$, which will lead to a bound on $|D_x\mathbf{b}_n(t,\cdot)|(\Omega)$. Pick a point $x_0 \in \Omega$. From (3.7), (3.8) and (3.10), the function $\mathbf{d}_n(t,\cdot) := \dfrac{\cdot - x_0}{t} - \mathbf{b}_n(t,\cdot)$ is in $BV_{\mathrm{loc}}(\mathbb{R}^d)$ and

$$\mathrm{div}_x \mathbf{d}_n(t,x) = \frac{1}{2t}\cdot\sum_{i\neq j} \langle y_i - y_j, \nu_i(x)\rangle\, \mathcal{H}^{d-1}_{\llcorner \partial\mathcal{V}_i\cap\partial\mathcal{V}_j}$$

is a positive Radon measure. In particular, this implies that

$$|\mathrm{div}_x \mathbf{d}_n(t,\cdot)|(\Omega) = \int_\Omega \mathrm{div}_x \mathbf{d}_n(t,\cdot)\, dx.$$

Let $\rho_\varepsilon \in \mathcal{C}^\infty_c(\mathbb{R}^d)$ be a family of mollifiers, i.e., $\rho_\varepsilon(x) = \varepsilon^{-d}\rho\left(\frac{x}{\varepsilon}\right)$ for $\rho \in \mathcal{C}^\infty_c(\mathbb{R}^d)$ satisfying $\rho(x) \geq 0$, $\rho(x) = \rho(-x)$, $\mathrm{supp}(\rho) \subset B_d(0,1)$ and $\int_{\mathbb{R}^d} \rho(x)dx = 1$. For every test function $\varphi_\varepsilon = \chi_\Omega * \rho_\varepsilon$, it holds that

$$\int_{\mathbb{R}^d} \varphi_\varepsilon \mathrm{div}_x \mathbf{d}_n(t,\cdot)\, dx = -\int_{\mathbb{R}^d} \mathbf{d}_n(t,x)\cdot\nabla\varphi_\varepsilon(x)dx \leq \|\mathbf{d}_n(t,\cdot)\|_{\mathbf{L}^\infty(\mathbb{R}^d)}\cdot\int_{\mathbb{R}^d} |\nabla\varphi_\varepsilon(x)|dx.$$

Thus, taking $\varepsilon \to 0+$, we get

$$\int_\Omega \mathrm{div}_x \mathbf{d}_n(t,\cdot)\, dx \;\leq\; \|\mathbf{d}_n(t,\cdot)\|_{\mathbf{L}^\infty(\mathbb{R}^d)}\cdot\mathcal{H}^{d-1}(\partial\Omega)$$

and (3.11) yields

$$|D_x\mathbf{b}_n(t,\cdot)|(\Omega) \;\le\; \frac{1}{\lambda_{\beta_n}}\cdot\left(\|\mathbf{b}_n(t,\cdot)\|_{\mathbf{L}^\infty(\mathbb{R}^d)}+\frac{\mathrm{diam}(\Omega)}{t}\right)\cdot\mathcal{H}^{d-1}(\partial\Omega)+\frac{\sqrt{d}}{t}\cdot|\Omega|. \tag{3.12}$$

Step 3. Finally, to achieve (3.6) by taking $n\to\infty$ in (3.12), we first claim that $\mathbf{b}_n(t,\cdot)$ converges to $\mathbf{b}(t,\cdot)$ in $\mathbf{L}^1_{\mathrm{loc}}$. Since the sequence $\mathbf{b}_n(t,\cdot)$ is bounded in $\mathbf{L}^\infty(\mathbb{R}^d)$ and the set $\left(\bigcup_{n\ge1}\Sigma_t^n\bigcup\Sigma_t\right)$ has zero Lebesgue measure with $\Sigma_t=\{x\in\mathbb{R}^d:\#\mathbf{b}(t,x)\ge2\}$, it is sufficient to show that

$$\lim_{n\to\infty}\mathbf{b}_n(t,x) \;=\; \mathbf{b}(t,x) \qquad \text{for all } x\in\mathbb{R}^d\backslash\left(\bigcup_{n\ge1}\Sigma_t^n\bigcup\Sigma_t\right).$$

Given $x\in\mathbb{R}^d\backslash\left(\bigcup_{n\ge1}\Sigma_t^n\bigcup\Sigma_t\right)$, assume by a contradiction that there exists a subsequence $\mathbf{b}_{n_k}(t,x)$ converging to some $w\neq\mathbf{b}(t,x)$. From Lemma 3.1, we have

$$\begin{aligned}u(t,x) \;&=\; \lim_{n_k\to\infty}u_{n_k}(t,x) \;=\; \lim_{n_k\to\infty}u_0(x-t\mathbf{b}_{n_k}(t,x))+t\cdot L(\mathbf{b}_{n_k}(t,x))\\ &=\; u_0(x-tw)+t\cdot L(w) \;=\; u_0(x-tw)+t\cdot L\left(\frac{x-(x-tw)}{t}\right).\end{aligned}$$

Thus, $\mathbf{b}(t,x)$ is not a singleton and this yields a contradiction. By [2, Proposition 3.13], (3.12) and Lemma 3.1, the function $\mathbf{b}_n(t,\cdot)$ converges weakly to $\mathbf{b}(t,\cdot)$ in $BV(\Omega,\mathbb{R}^d)$ with

$$\begin{aligned}|D_x\mathbf{b}(t,\cdot)|(\Omega) \;&\le\; \liminf_{n\to\infty}|D_x\mathbf{b}_n(t,\cdot)|(\Omega)\\ &\le\; \left(\frac{1}{\limsup\limits_{n\to\infty}\lambda_{\beta_n}}\right)\cdot\left(\Lambda_M+\frac{\mathrm{diam}(\Omega)}{t}\right)\cdot\mathcal{H}^{d-1}(\partial\Omega)+\frac{\sqrt{d}}{t}\cdot|\Omega|\end{aligned}$$

and this yields (3.6). □

As a direct consequence of Theorem 3.2, the following holds.

Corollary 3.3 *The map $S_T:\mathbf{Lip}\left(\mathbb{R}^d\right)\to\mathbf{Lip}\left(\mathbb{R}^d\right)$ is compact in $\mathbf{W}^{1,1}_{\mathrm{loc}}\left(\mathbb{R}^d\right)$ for every time $T>0$.*

Proof. Given

$$\mathcal{U}_{[m,M]} \;:=\; \left\{\bar{u}\in\mathbf{Lip}(\mathbb{R}^d):|\bar{u}(0)|\le m,\mathrm{Lip}[\bar{u}]\le M\right\}$$

for any sequence of initial data $(\bar{u}_n)_{n\ge1}\subseteq\mathcal{U}_{[m,M]}$, we set

$$v_n(x) \;:=\; S_T(\bar{u}_n)(x) \qquad \text{for all } x\in\mathbb{R}^d, n\ge1.$$

From Theorem 3.2, it holds

$$\|DH(Dv_n)\|_{\mathbf{L}^\infty} \;\le\; \Lambda_M \qquad \text{and} \qquad DH(Dv_n)\in BV_{\mathrm{loc}}(\mathbb{R}^d).$$

and

$$\sup_{|x|\le R}v(x) \;\le\; m+MR+T\cdot\sup_{|p|\le\Lambda_M}L\left(q\right) \qquad \text{for all } R>0.$$

By Helly's theorem, there exists a subsequence $(v_{n_k})_{k\ge1}$ and $w\in BV_{\mathrm{loc}}(\mathbb{R}^d)$ such that

- $v_{n_k}(0)$ converges to some $\bar{v}_0 \in \mathbb{R}$;
- $DH(Dv_{n_k})$ converges to w point-wise and

$$\lim_{k\to\infty} \|DH(Dv_{n_k}) - w\|_{\mathbf{L}^1(B_d(0,R))} \;=\; 0 \qquad \text{for all } R > 0.$$

This implies that $Dv_{n_k} = DL(DH(Dv_{n_k}))$ converges to $DL(w)$ point-wise and

$$\lim_{n_k\to+\infty} \|Dv_{n_k} - DL(w)\|_{\mathbf{L}^1(B_d(0,R))} \;=\; 0\,.$$

Thus, denote by

$$\bar{v}^R_{n_k} \;:=\; \frac{1}{|B_d(0,R)|}\cdot\int_{B_d(0,R)} v_{n_k}(x)dx$$

the average of v_{n_k} in $B_d(0,R)$, we have

$$\begin{aligned}\lim_{n_k\to\infty}\left(\bar{v}^R_{n_k} - v_{n_k}(0)\right) \;&=\; \lim_{n_k\to\infty}\frac{1}{|B_d(0,R)|}\cdot\int_0^1\int_{B_d(0,R)} Dv_{n_k}(sx)(x)dxds\\ &=\; \frac{1}{|B_d(0,R)|}\cdot\int_0^1\int_{B_d(0,R)} DL(w(sx))dxds \;:=\; \bar{v}^R\end{aligned}$$

and this yields $\lim_{n_k\to\infty}\bar{v}^R_{n_k} = \bar{v}_0 + \bar{v}^R$. On the other hand, by the Poincaré inequality, it holds that

$$\left(\left\|\left(v_{n_k} - \bar{v}^R_{n_k}\right) - \left(v_{n'_k} - \bar{v}^R_{n'_k}\right)\right\|_{\mathbf{L}^1(B_d(0,R))}\right) \;\leq\; R\cdot\left\|Dv_{n_k} - Dv_{n'_k}\right\|_{\mathbf{L}^1(0,R)}.$$

Therefore, the sequence $(v_{n_k})_{k\geq1}$ is a Cauchy sequence in $\mathbf{W}^{1,1}(B_d(0,R))$ for every $R>0$ and thus converges to $\bar{v}$ in $\mathbf{W}^{1,1}_{\mathrm{loc}}(\mathbb{R}^d)$. □

4 Metric entropy in $\mathbf{W}^{1,1}$ for S_T

In this section, we shall quantify the degree of compactness of the map $S_T : \mathbf{Lip}\left(\mathbb{R}^d\right) \to \mathbf{Lip}\left(\mathbb{R}^d\right)$ for a given time $T>0$. More precisely, given constants $m, M, R > 0$, considering the set of initial data

$$\mathcal{U}_{[m,M]} = \left\{\bar{u}\in\mathbf{Lip}(\mathbb{R}^d) : |\bar{u}(0)|\leq m, \mathrm{Lip}[\bar{u}]\leq M\right\}, \tag{4.1}$$

we establish upper and lower estimates for the ε-entropy of the following restricted solution set at time T in $\mathbf{W}^{1,1}(\square_R)$

$$S_{T,R}\left(\mathcal{U}_{[m,M]}\right) := \left\{v_{\llcorner\square_R} \;:\; v\in S_T\left(\mathcal{U}_{[m,M]}\right)\right\}. \tag{4.2}$$

In order to do so, let us introduce continuous real-valued functions Ψ_M, Φ_M defined on $[0,M]$ for $M>0$ such that $\Psi_M(0) = \Phi_M(0) = 0$ and

$$\begin{cases}\Psi_M(s) &= s\cdot\displaystyle\min_{p,q\in\overline{B}_d(0,M),|p-q|\geq s}\frac{|DH(p)-DH(q)|}{|p-q|}\\ \Phi_M(s) &= s\cdot\displaystyle\min_{p\in\overline{B}_d\left(0,M-\frac{s}{2}\right)}\left(\max_{q\in\overline{B}_d(p,\frac{s}{2})}\left\|D^2H(q)\right\|_\infty\right)\end{cases} \quad \text{for all } s\in(0,M]. \tag{4.3}$$

Here, $\left|D^2H(q)(v)\right|$ is the matrix norm of $D^2H(q)(v)$ and $\left\|D^2H(q)\right\|_\infty := \max_{|v|\leq 1}\left|D^2H(q)(v)\right|$. Notice that both maps $s \mapsto \Psi_M(s)$ and $s \mapsto \Phi_M(s)$ are strictly increasing. Moreover, the strict convexity of H implies that

$$0 < \Psi_M(s) \leq \Phi_M(s) \leq M \cdot \max_{p\in \overline{B}_d(0,M)} \left\|D^2H(p)\right\|_\infty \quad \text{for all } s \in (0, M].$$

For convenience, we now rewrite Theorem 1.3 as our second main theorem.

Theorem 4.1 *Assume that $H \in \mathcal{C}^2(\mathbb{R}^d)$ and satisfies* **(H1)**-**(H2)**. *Then for every $\varepsilon > 0$ sufficiently small, it holds that*

$$C_1 \cdot \left(\Phi_M\left(\frac{\varepsilon}{R_1}\right)\right)^{-d} \leq \mathcal{H}_\varepsilon\left(S_{T,R}(\mathcal{U}_{[m,M]})\Big|\mathbf{W}^{1,1}(\square_R)\right) \ \leq \ C_2 \cdot \left(\Psi_M\left(\frac{\varepsilon}{R_2}\right)\right)^{-d}. \quad (4.4)$$

for some constants $C_1, C_2, R_1, R_2 > 0$ depending only on $m, M, R, T > 0$.

Before proving Theorem 4.1 in the next two subsections, we present some cases in which the estimates in (4.4) are sharp.

Remark 4.2 *Given $\alpha > 1$, the Hamiltonian $H(p) = |p|^{2\alpha}$ is not strongly convex but satisfies* **(H1)**-**(H2)**. *In this case, we compute for all $v, p \in \mathbb{R}^d$ that*

$$\begin{aligned} 2\alpha|p|^{2(\alpha-1)}|v|^2 \ \leq \ \left\langle DH^2(p)(v), v\right\rangle \ &= \ 2\alpha|p|^{2(\alpha-1)}|v|^2 + 4\alpha(\alpha-1)|p|^{2(\alpha-2)}|\langle p, v\rangle|^2 \\ &\leq \ 2\alpha(2\alpha-1)|p|^{2(\alpha-1)}|v|^2. \end{aligned}$$

Thus, there exist constants $\alpha_1, \alpha_2 > 0$ such that

$$\alpha_1 s^{2\alpha-1} \leq \Psi_M(s) \leq \Phi_M(s) \leq \alpha_2 s^{2\alpha-1} \qquad \text{for all } s \in [0, M],$$

and (4.4) yields

$$\mathcal{H}_\varepsilon\left(S_{T,R}(\mathcal{U}_{[m,M]})\big|\mathbf{W}^{1,1}(\square_R)\right) \approx \varepsilon^{-(2\alpha-1)d}.$$

Remark 4.3 *If $H \in \mathcal{C}^2(\mathbb{R}^d)$ is strongly convex then H satisfies* **(H1)**-**(H2)** *and*

$$\alpha_1 s < \Psi_M(s) \leq \Phi_M(s) < \alpha_2 s \qquad \text{for all } s \in [0, M],$$

for some $\alpha_1, \alpha_2 > 0$. Thus, (4.4) yields the same result as in [3] that

$$\mathcal{H}_\varepsilon\left(S_{T,R}(\mathcal{U}_{[m,M]})\big|\mathbf{W}^{1,1}(\square_R)\right) \approx \varepsilon^{-d}$$

for every $\varepsilon > 0$ sufficiently small.

In the one-dimensional case, from Theorem 4.1 we can obtain an estimate similar to that established in [8, Remark 1.4] for scalar conservation laws with strictly convex fluxes.

Remark 4.4 *For $d = 1$, every strictly convex $H \in \mathcal{C}^2(\mathbb{R})$ satisfies (1.3). In addition, assume that H has polynomial degeneracy, i.e., the set $I_H = \{\omega \in \mathbb{R} : H''(\omega) = 0\} \neq \varnothing$ is finite and for each $w \in I_H$, there exists a natural number $p_\omega \geq 2$ such that*

$$H^{(p_\omega+1)}(\omega) \neq 0 \quad \text{and} \quad H^{(j)}(\omega) = 0 \quad \text{for all } j \in \{2, \dots, p_\omega\}.$$

The polynomial degeneracy of H is defined by $\mathbf{p}_H := \max_{\omega \in I_H} p_\omega$. For every $M > \operatorname{argmax}_{w \in I_H} p_\omega$, there exist constants $\alpha_1, \alpha_2 > 0$ such that

$$\alpha_1 \cdot s^{\mathbf{p}_H} < \Psi_M(s) \leq \Phi_M(s) < \alpha_2 \cdot s^{\mathbf{p}_H} \qquad \text{for all } s \in [0, M].$$

Thus, (4.4) implies that

$$\mathcal{H}_\varepsilon\left(S_{T,R}(\mathcal{U}_{[m,M]})\middle|\mathbf{W}^{1,1}(\square_R)\right) \approx \varepsilon^{-\mathbf{p}_H}$$

for every $\varepsilon > 0$ sufficiently small.

4.1 Upper estimate of $\mathcal{H}_\varepsilon\left(S_{T,R}(\mathcal{U}_{[m,M]})\middle|\mathbf{W}^{1,1}(\square_R)\right)$

Towards the upper estimate of $\mathcal{H}_\varepsilon\left(S_{T,R}(\mathcal{U}_{[m,M]})\middle|\mathbf{W}^{1,1}(\square_R)\right)$ in (4.4), we first provide a bound on the $\mathbf{L}^1$-distance between elements Du_1 and Du_2 in terms of the $\mathbf{L}^1$-distance between $DH(Du_1)$ and $DH(Du_2)$ for every $u_1, u_2 \in S_{T,R}(\mathcal{U}_{[m,M]})$ by using the function Ψ_M defined in (4.3). Observing that the map $s \mapsto \dfrac{\Psi_M(s)}{s}$ is monotone increasing and

$$\Psi_M(|p-q|) \leq |DH(p) - DH(q)| \quad \text{for all } p, q \in \overline{B}_d(0, M), \tag{4.5}$$

we prove the following lemma.

Lemma 4.5 *For any $u_1, u_2 \in S_{T,R}(\mathcal{U}_{[m,M]})$, it holds that*

$$\|Du_1 - Du_2\|_{\mathbf{L}^1(\square_R)} \leq \left(2^d R^d + 1\right) \cdot \Psi_M^{-1}\left(\|\mathbf{b}_1 - \mathbf{b}_2\|_{\mathbf{L}^1(\square_R)}\right) \tag{4.6}$$

with $\mathbf{b}_1 := DH(Du_1)$ and $\mathbf{b}_2 := DH(Du_2)$.

Proof. For simplicity, setting $\alpha := \Psi_M^{-1}\left(\|\mathbf{b}_1 - \mathbf{b}_2\|_{\mathbf{L}^1(\square_R)}\right)$, we claim that

$$|Du_1(x) - Du_2(x)| \leq \alpha \cdot \max\left\{1, \frac{|\mathbf{b}_1(x) - \mathbf{b}_2(x)|}{\|\mathbf{b}_1 - \mathbf{b}_2\|_{\mathbf{L}^1(\square_R)}}\right\} \qquad \text{for } a.e.\ x \in \square_R. \tag{4.7}$$

Indeed, assume that $|Du_1(x) - Du_2(x)| > \alpha$. From (4.5), it holds that

$$\begin{aligned} |Du_1(x) - Du_2(x)| &= \frac{|Du_1(x) - Du_2(x)|}{|DH(Du_1(x)) - DH(Du_2(x))|} \cdot |\mathbf{b}_1(x) - \mathbf{b}_2(x)| \\ &\leq \frac{|Du_1(x) - Du_2(x)|}{\Psi_M(|Du_1(x) - Du_2(x)|)} \cdot |\mathbf{b}_1(x) - \mathbf{b}_2(x)|. \end{aligned}$$

By the monotone increasing property of the map $s \mapsto \Psi_M(s)/s$, one has

$$|Du_1(x) - Du_2(x)| \le \frac{\alpha}{|\Psi_M(\alpha)|} \cdot |\mathbf{b}_1(x) - \mathbf{b}_2(x)| \;=\; \alpha \cdot \frac{|\mathbf{b}_1(x) - \mathbf{b}_2(x)|}{\|\mathbf{b}_1 - \mathbf{b}_2\|_{\mathbf{L}^1(\Box_R)}}$$

and this implies (4.7). Hence, the $\mathbf{L}^1$-distance between Du_1 and Du_2 is bounded by

$$\begin{aligned}
\|Du_1 - Du_2\|_{\mathbf{L}^1(\Box_R)} &= \int_{\Box_R} |Du_1(x) - Du_2(x)| dx \\
&\le \alpha \cdot \int_{\Box_R} \left(1 + \frac{|\mathbf{b}_1(x) - \mathbf{b}_2(x)|}{\|\mathbf{b}_1 - \mathbf{b}_2\|_{\mathbf{L}^1(\Box_R)}}\right) dx = (2^d R^d + 1)\alpha \\
&= \left(2^d R^d + 1\right) \cdot \Psi_M^{-1}\left(\|\mathbf{b}_1 - \mathbf{b}_2\|_{\mathbf{L}^1(\Box_R)}\right) \qquad (4.8)
\end{aligned}$$

and this yields (4.6). $\Box$

Proof of the upper estimate of $\mathcal{H}_\varepsilon\left(S_{T,R}(\mathcal{U}_{[m,M]}) \middle| \mathbf{W}^{1,1}(\Box_R)\right)$ in Theorem 4.1

1. From Theorem 1.2, for any $v \in S_{T,R}(\mathcal{U}_{[m,M]})$, one has

$$|DH(Dv)|(\Box_R) \le V_T \qquad \text{and} \qquad \|v\|_{\mathbf{L}^\infty|(\Box_R)} \le m_T \qquad (4.9)$$

with $V_T := \frac{d2^dR^{d-1}}{\gamma_M} \cdot \left(\Lambda_M + \frac{2\sqrt{d}R}{T}\right) + \frac{\sqrt{d}2^dR^d}{T}$ and $m_T := m + \sqrt{d}MR + T \cdot \sup_{|q| \le \Lambda_M} L(q)$. In particular, the average value $\bar{v}^R$ of v satisfies

$$\bar{v}^R = \frac{1}{\text{Vol}(\Box_R)} \cdot \int_{\Box_R} v(x)\ dx \;\in\; [-m_T, m_T].$$

Given $\varepsilon' > 0$, we cover the interval $[-m_T, m_T]$ by $K_{\varepsilon'} = \left\lfloor \frac{m_T}{\Psi_M^{-1}(\varepsilon')} \right\rfloor + 1$ small intervals with length $2\Psi_M^{-1}(\varepsilon')$ such that

$$[-m_T, m_T] \subseteq \bigcup_{i=1}^{K_{\varepsilon'}} B\big(a_i, \Psi_M^{-1}(\varepsilon')\big) \quad \text{for some } a_i \in [-m_T, m_T]$$

and then decompose the set $S_{T,R}\big(\mathcal{U}_{[m,M]}\big)$ into $K_{\varepsilon'}$ subsets as follows:

$$S_{T,R}\left(\mathcal{U}_{[m,M]}\right) \subseteq \bigcup_{i=1}^{K_{\varepsilon'}} \mathcal{A}_i, \qquad \mathcal{A}_i := \left\{v \in S_{T,R}\left(\mathcal{U}_{[m,M]}\right) : \bar{v}^R \in B\left(a_i, \Psi_M^{-1}(\varepsilon')\right)\right\}.$$

Recalling Definition 1.1, we have

$$\mathcal{N}_\varepsilon\left(S_{T,R}\left(\mathcal{U}_{[m,M]}\right) \middle| \mathbf{W}^{1,1}(\Box_R)\right) \le \sum_{i=1}^{K_{\varepsilon'}} \mathcal{N}_\varepsilon\left(\mathcal{A}_i \middle| \mathbf{W}^{1,1}(\Box_R)\right) \qquad (4.10)$$

for all $\varepsilon > 0$.

2. For a given $i \in \{1, 2, \ldots, K_{\varepsilon'}\}$, we are going to provide an upper bound on the covering number $\mathcal{N}_\varepsilon\left(\mathcal{A}_i \middle| \mathbf{W}^{1,1}(\Box_R)\right)$ by introducing the set $\mathcal{B}_i := \{DH(Dv) : v \in \mathcal{A}_i\}$. From (2.2) and (4.9), one has that $\mathcal{B}_i \subseteq \mathcal{F}_{[R,m_T,V_T]}$ with $\mathcal{F}_{[R,m_T,V_T]}$ defined as in (2.2). By Corollary 2.5, if $\varepsilon' > 0$ is sufficiently small then it holds that

$$\mathcal{H}_{\varepsilon'/2}\left(\mathcal{B}_i \middle| \mathbf{L}^1(\Box_R)\right) \leq \Gamma^+ \cdot (\varepsilon')^{-d}, \qquad \Gamma^+ = 48\sqrt{d} \cdot \left(12d\sqrt{d}RV_T\right)^d. \tag{4.11}$$

By the definition of $\mathcal{H}_{\varepsilon'/2}\left(\mathcal{B}_i \middle| \mathbf{L}^1(\Box_R)\right)$, there exists a set $\left\{\mathbf{v}_1, \ldots, \mathbf{v}_{\beta_{\varepsilon'}}\right\} \subset \mathcal{A}_i$ with $\beta_{\varepsilon'} \leq 2^{\Gamma^+ \cdot (\varepsilon')^{-d}}$ such that

$$\mathcal{B}_i \subseteq \bigcup_{j=1}^{\beta_{\varepsilon'}} B_{\mathbf{L}^1}(\mathbf{b}_j, \varepsilon'), \qquad \mathbf{b}_j \ := \ DH(D\mathbf{v}_j).$$

In particular, for any given $v \in \mathcal{A}_i$, it holds that

$$\|DH(Dv) - \mathbf{b}_{j_0}\|_{\mathbf{L}^1(\Box_R)} \ < \ \varepsilon' \quad \text{for some } j_0 \in \overline{1, \beta_{\varepsilon'}}.$$

Recalling Lemma 4.5, we obtain that

$$\begin{aligned}
\|Dv - D\mathbf{v}_{j_0}\|_{\mathbf{L}^1(\Box_R)} \ &\leq \ \left(2^d R^d + 1\right) \cdot \Psi_M^{-1}\left(\|DH(Dv) - \mathbf{b}_{j_0}\|_{\mathbf{L}^1(\Box_R)}\right) \\
&\leq \ \left(2^d R^d + 1\right) \cdot \Psi_M^{-1}(\varepsilon')
\end{aligned}$$

and the Poincaré inequality in Theorem 2.4 yields

$$\begin{aligned}
\left\|\left(v - \bar{v}^R\right) - \left(\mathbf{v}_{j_0} - \bar{\mathbf{v}}_{j_0}^R\right)\right\|_{\mathbf{L}^1(\Box_R)} \ &\leq \ \sqrt{d}R \cdot \|Dv - D\mathbf{v}_{j_0}\|_{\mathbf{L}^1(\Box_R)} \\
&\leq \ \sqrt{d}R\left(2^d R^d + 1\right) \cdot \Psi_M^{-1}(\varepsilon').
\end{aligned}$$

On the other hand, since $v, \mathbf{v}_{j_0} \in \mathcal{A}_i$, one has

$$\left|\bar{v}^R - \bar{\mathbf{v}}_{j_0}^R\right| \leq \left|\bar{v}^R - a_i\right| + \left|\bar{\mathbf{v}}_{j_0}^R - a_i\right| \ \leq \ 2\Psi_M^{-1}(\varepsilon').$$

Thus, the $\mathbf{W}^{1,1}$-distance between v and $\mathbf{v}_{j_0}$ can be estimated by

$$\begin{aligned}
\|v - \mathbf{v}_{j_0}\|_{\mathbf{W}^{1,1}(\Box_R)} \ &\leq \ \left|\bar{v}^R - \bar{\mathbf{v}}_{j_0}^R\right| \cdot \left|\Box_R\right| + \|Dv - D\mathbf{v}_{j_0}\|_{\mathbf{L}^1(\Box_R)} \\
&+ \ \left\|\left(v - \bar{v}^R\right) - \left(\mathbf{v}_{j_0} - \bar{\mathbf{v}}_{j_0}^R\right)\right\|_{\mathbf{L}^1(\Box_R)} \leq R^+ \cdot \Psi_M^{-1}(\varepsilon')
\end{aligned}$$

with $R^+ := \left(2^d R^d + 1\right)(3 + \sqrt{d}R)$. Finally, by choosing $\varepsilon' = \Psi_M\left(\dfrac{\varepsilon}{R^+}\right)$, we have that $\mathcal{A}_i \subseteq \bigcup_{i=1}^{\beta_{\varepsilon'}} B_{\mathbf{W}^{1,1}}(\mathbf{v}_i, \varepsilon)$ and

$$\mathcal{N}_\varepsilon\left(\mathcal{A}_i \middle| \mathbf{W}^{1,1}(\Box_R)\right) \leq \beta_{\varepsilon'} \ = \ 2^{\Gamma^+ \cdot \left(\Psi_M\left(\frac{\varepsilon}{R^+}\right)\right)^{-d}}.$$

Therefore, from (4.10), one gets

$$\mathcal{N}_\varepsilon\left(S_{T,R}(\mathcal{U}_{[m,M]})\Big|\mathbf{W}^{1,1}(\Box_R)\right) \leq \left(\left\lfloor\frac{m_T R^+}{\varepsilon}\right\rfloor + 1\right)\cdot 2^{\Gamma^+\cdot\left(\Psi_M\left(\frac{\varepsilon}{R^+}\right)\right)^{-d}}$$

and this yields the second inequality in (4.4) for $\varepsilon > 0$ sufficiently small. $\Box$

4.2 Lower estimate of $\mathcal{H}_\varepsilon\left(S_{T,R}(\mathcal{U}_{[m,M]})\big|\mathbf{W}^{1,1}(\Box_R)\right)$

In this subsection, we shall prove the first inequality in (4.4). In order to do so, for any given $p \in \mathbb{R}^d$, let $\Phi(\cdot,p) : [0,\infty) \to [0,\infty)$ be the strictly increasing continuous function defined by $\Phi(0,p) = 0$ and

$$\Phi(s,p) = s\cdot\left(\max_{p'\in\overline{B}_d(p,\frac{s}{2})}\left\|D^2H(p')\right\|_\infty\right), \qquad s > 0.$$

From the definition of Φ_M in (4.3), it holds that

$$\Phi_M(s) = \min_{p\in\overline{B}_d\left(0,M-\frac{s}{2}\right)}\Phi(s,p), \qquad s\in[0,M]. \tag{4.12}$$

Let us recall the constant in the assumption **(H2)**

$$\lambda_r = \inf_{p\neq q\in\overline{B}(0,r)}\left\langle\frac{DH(p)-DH(q)}{|DH(p)-DH(q)|},\frac{p-q}{|p-q|}\right\rangle > 0, \qquad r > 0. \tag{4.13}$$

The following proposition shows that a solution to (1.1) with a semiconvex initial condition preserves the semiconvexity on a given time interval, provided the semiconvexity constant of the initial data is sufficiently small in absolute value.

Proposition 4.1 *Assume that $H \in \mathcal{C}^2(\mathbb{R}^d)$ and satisfies* **(H1)**-**(H2)**. *Given $T, M, r > 0$ and $\bar{p} \in \overline{B}_d\left(0, M - \frac{r}{2}\right)$, let $\bar{u}$ be a semiconvex function with semiconvexity constant $-K$ such that*

$$D^-\bar{u}(\mathbb{R}^d) \subseteq \overline{B}_d\left(\bar{p},\frac{r}{2}\right), \qquad K \leq \frac{\lambda_M}{4T}\cdot\frac{r}{\Phi(r,\bar{p})}. \tag{4.14}$$

Then, the map $(t,x)\mapsto S_t(\bar{u})(x)$ is a classical solution for $0 < t \leq T$ and

$$DS_t(\bar{u})(x) \in \overline{B}_d\left(\bar{p},\frac{r}{2}\right) \qquad \text{for all } (t,x)\in(0,T]\times\mathbb{R}^d.$$

Proof. For simplicity, we set

$$u(t,x) := S_t(\bar{u})(x) \qquad \text{for all } (t,x)\in[0,\infty)\times\mathbb{R}^d.$$

It is well-known from [9, Theorem 5.3.8] that $u(t,\cdot)$ is locally semiconcave for every $t > 0$. Thus, by Proposition 2.2, it is sufficient to show that $u(t,\cdot)$ is semiconvex with some semiconvexity constant $-C < 0$ for all $t \in [0,T]$, i.e., for any fixed $(t,x) \in [0,T)\times\mathbb{R}^d$, it holds that

$$u(t,x+h) + u(t,x-h) - 2u(t,x) \geq -C\cdot|h|^2 \quad \text{for all } h\in\mathbb{R}^d. \tag{4.15}$$

By the Lipschitz continuity of $u(t,\cdot)$, we can assume that $u(t,\cdot)$ is differentiable at $x \pm h$. In this case, $\mathbf{b}(t, x \pm h)$ reduce to a single value denoted by $\mathbf{b}^{\pm} = DH(\mathbf{p}^{\pm})$ with $\mathbf{p}^{\pm} = D_x u(t, x \pm h)$. Moreover, since $x \pm h - t\mathbf{b}^{\pm}$ is a minimizer in $\min_{y\in\mathbb{R}^d} \left\{ \bar{u}(y) + t \cdot L\left(\frac{x\pm h - y}{t}\right)\right\}$, it holds that

$$\begin{cases} \mathbf{p}^{\pm} \in D^{-}\bar{u}(x \pm h - t\mathbf{b}^{\pm}) \subseteq \overline{B}_d\left(\bar{p}, \dfrac{r}{2}\right) \subseteq \overline{B}_d(0, M), \\ \\ u(t, x\pm h) = \bar{u}(x \pm h - t\mathbf{b}^{\pm}) + t\cdot L(\mathbf{b}^{\pm}). \end{cases} \tag{4.16}$$

Since $\bar{u}$ is semiconvex with semiconvexity constant $-K$, denoting $x^{\pm} := x \pm h$, from (iii) of Proposition 2.1 one can get that

$$\left\langle \mathbf{p}^{+} - \mathbf{p}^{-}, x^{+} - x^{-} - t\left(\mathbf{b}^{+} - \mathbf{b}^{-}\right)\right\rangle \geq -K \cdot \left|2h - t\left(\mathbf{b}^{+} - \mathbf{b}^{-}\right)\right|^2$$

and

$$\begin{aligned} \left\langle \mathbf{p}^{+} - \mathbf{p}^{-}, \mathbf{b}^{+} - \mathbf{b}^{-}\right\rangle &\leq \frac{K}{t}\cdot\left|2h - t\left(\mathbf{b}^{+} - \mathbf{b}^{-}\right)\right|^2 + \frac{2|h|}{t}\cdot|\mathbf{p}^{+} - \mathbf{p}^{-}| \\ &\leq 2Kt|\mathbf{b}^{+} - \mathbf{b}^{-}|^2 + \frac{8K|h|^2}{t} + \frac{2|h|}{t}\cdot|\mathbf{p}^{+} - \mathbf{p}^{-}| \\ &\leq 2KT|DH(\mathbf{p}^{+}) - DH(\mathbf{p}^{-})|^2 + \frac{8K|h|^2}{t} + \frac{2|h|}{t}\cdot|\mathbf{p}^{+} - \mathbf{p}^{-}|. \end{aligned}$$

Since $\mathbf{p}^{\pm} \in \overline{B}_d(\bar{p}, \frac{r}{2})$, it holds that

$$|DH(\mathbf{p}^{+}) - DH(\mathbf{p}^{-})| \leq \frac{\Phi(r,\bar{p})}{r}\cdot|\mathbf{p}^{+} - \mathbf{p}^{-}|.$$

Thus, recalling

$$\lambda_M = \inf_{p\neq q\in\overline{B}(0,M)} \left\langle \frac{DH(p) - DH(q)}{|DH(p) - DH(q)|}, \frac{p-q}{|p-q|}\right\rangle > 0, \qquad M > 0,$$

using (4.14) and (4.16), we estimate

$$\begin{aligned} 2KT|DH(\mathbf{p}^{+}) - DH(\mathbf{p}^{-})|^2 &\leq 2KT\cdot\frac{\Phi(r,\bar{p})}{r}\cdot|DH(\mathbf{p}^{+}) - DH(\mathbf{p}^{-})|\cdot|\mathbf{p}^{+} - \mathbf{p}^{-}| \\ &\leq \frac{\lambda_M}{2}\cdot|DH(\mathbf{p}^{+}) - DH(\mathbf{p}^{-})|\cdot|\mathbf{p}^{+} - \mathbf{p}^{-}| = \frac{\lambda_M}{2}\cdot|\mathbf{b}^{+} - \mathbf{b}^{-}|\cdot|\mathbf{p}^{+} - \mathbf{p}^{-}| \end{aligned}$$

and

$$\left\langle \mathbf{p}^{+} - \mathbf{p}^{-}, \mathbf{b}^{+} - \mathbf{b}^{-}\right\rangle \leq \frac{\lambda_M}{2}\cdot|\mathbf{b}^{+} - \mathbf{b}^{-}|\cdot|\mathbf{p}^{+} - \mathbf{p}^{-}| + \frac{8K|h|^2}{t} + \frac{2|h|}{t}\cdot|\mathbf{p}^{+} - \mathbf{p}^{-}|. \tag{4.17}$$

On the other hand, from (4.13) we deduce that

$$\begin{aligned} \left\langle \mathbf{p}^{+} - \mathbf{p}^{-}, \mathbf{b}^{+} - \mathbf{b}^{-}\right\rangle &= \left\langle \mathbf{p}^{+} - \mathbf{p}^{-}, DH(\mathbf{p}^{+}) - DH(\mathbf{p}^{-})\right\rangle \\ &\geq \lambda_M\cdot|DH(\mathbf{p}^{+}) - DH(\mathbf{p}^{-})|\cdot|\mathbf{p}^{+} - \mathbf{p}^{-}| = \lambda_M\cdot|\mathbf{b}^{+} - \mathbf{b}^{-}|\cdot|\mathbf{p}^{+} - \mathbf{p}^{-}| \end{aligned}$$

and (4.17) implies that

$$\frac{\lambda_M}{2}\cdot|t(\mathbf{b}^+-\mathbf{b}^-)|\cdot|\mathbf{p}^+-\mathbf{p}^-| \ \leq\ 8K|h|^2+2|h|\cdot|\mathbf{p}^+-\mathbf{p}^-|. \tag{4.18}$$

Observe that if $|\mathbf{p}^+-\mathbf{p}^-|\leq K|h|$ then

$$|\mathbf{b}^+-\mathbf{b}^-|=|DH(\mathbf{p}^+)-DH(\mathbf{p}^-)|\leq\frac{\Phi(r,\bar{p})}{r}\cdot|\mathbf{p}^+-\mathbf{p}^-|\leq\frac{K\Phi(r,\bar{p})}{r}\cdot|h|.$$

Otherwise, (4.18) implies that $\frac{\lambda_M}{2}\cdot|t(\mathbf{b}^+-\mathbf{b}^-)|\leq 10|h|$. Hence, it holds that

$$|t(\mathbf{b}^+-\mathbf{b}^-)| \ \leq\ \left(\frac{KT\Phi(r,\bar{p})}{r}+\frac{20}{\lambda_M}\right)\cdot|h|. \tag{4.19}$$

By the Hopf-Lax representation formula, we have

$$\begin{aligned} u(t,x\pm h) &\ =\ \bar{u}(x\pm h-t\mathbf{b}^{\pm})+t\cdot L(\mathbf{b}^{\pm}) \\ u(t,x) &\ \leq\ \bar{u}\left(x-t\cdot\frac{\mathbf{b}^++\mathbf{b}^-}{2}\right)+t\cdot L\left(\frac{\mathbf{b}^++\mathbf{b}^-}{2}\right). \end{aligned}$$

Using the convexity of L and semiconvexity of $\bar{u}$, we estimate

$$\begin{aligned} u(t,x+h)+u(t,x-h)-2u(t,x) &\geq t\cdot\left[L(\mathbf{b}^+)+L(\mathbf{b}^-)-2L\left(\frac{\mathbf{b}^++\mathbf{b}^-}{2}\right)\right] \\ &\quad+\bar{u}(x+h-t\mathbf{b}^+)+\bar{u}(x-h-t\mathbf{b}^-)-2\bar{u}\left(x-t\cdot\frac{\mathbf{b}^++\mathbf{b}^-}{2}\right) \\ &\geq -K\cdot\left|2h-t\left(\mathbf{b}^+-\mathbf{b}^-\right)\right|^2\geq -8K|h|^2-2K|t(\mathbf{b}^+-\mathbf{b}^-)|^2 \\ &\geq -2K\cdot\left[4+\left(\frac{KT\Phi(r,\bar{p})}{r}+\frac{20}{\lambda_M}\right)^2\right]\cdot|h|^2 \end{aligned}$$

and this yields (4.15). □

Relying on the above Proposition and Corollary 2.2, we now proceed to prove the first inequality in (4.4).

Proof of the lower estimate of $\mathcal{H}_\varepsilon\left(S_{T,R}(\mathcal{U}_{[m,M]})\middle|\mathbf{W}^{1,1}(\Box_R)\right)$ in Theorem 4.1

1. Let us recall that $\mathcal{U}_{[m,M]}=\left\{\bar{u}\in\mathbf{Lip}(\mathbb{R}^d):|\bar{u}(0)|\leq m,\mathrm{Lip}[\bar{u}]\leq M\right\}$ and

$$\mathcal{SC}_{[r,K]}:=\left\{v\in\mathbf{Lip}(\mathbb{R}^d):\mathrm{Lip}[v]\leq r\text{ and }v\text{ is semiconcave with constant }K\right\}.$$

For any given $r>0$ and $p\in\mathbb{R}^d$, we denote by

$$\mathcal{W}_r^p:=\left\{\varphi=v+\langle p,\cdot\rangle:v\in\mathcal{SC}_{[\frac{r}{2},K_r]}\right\},\qquad K_r=\frac{\lambda_M}{4T}\cdot\frac{r}{\Phi_M(r)}.$$

From (4.12), there exists $p_r \in \overline{B}_d(0, M - \frac{r}{2})$ such that $\Phi(r, p_r) = \min_{p \in \overline{B}(0, M - \frac{r}{2})} \Phi(r, p)$. Now consider the operator $\mathcal{T} : \mathcal{W}_r^{p_r} \to \mathbf{Lip}(\mathbb{R}^d)$ such that for all $\varphi \in \mathcal{W}_r^p$, it holds that

$$\mathcal{T}(\varphi) \;=\; \varphi + S_T(\varphi_-)(0), \qquad \varphi_-(\cdot) := -\varphi(-\cdot). \tag{4.20}$$

By reversing the equation (1.1), we will show that

$$\mathcal{T}\left(\mathcal{W}_r^{p_r}\right) \subseteq S_T\left(\mathcal{U}_{[0,M]}\right). \tag{4.21}$$

Indeed, for a given $\varphi \in \mathcal{W}_r^{p_r}$, we define the following function

$$w_0(\cdot) := -\mathcal{T}(\varphi)(-\cdot) = \varphi_-(\cdot) - S_T(\varphi_-)(0).$$

Since $\varnothing \neq D^+\varphi(x) \subset \overline{B}_d\left(p_r, \frac{r}{2}\right)$ for all $x \in \mathbb{R}^d$, w_0 is semiconvex with a semiconvexity constant $-K_r$ and

$$D^- w_0(x) = p_r + D^+\varphi(-x) \subseteq \overline{B}_d\left(p_r, \frac{r}{2}\right) \qquad \text{for all } x \in \mathbb{R}^d.$$

Let $w(t, x) = S_t(w_0)(x)$ be the unique viscosity solution of (1.1) with initial datum $w(0, x) = w_0$. Recalling Proposition 4.1 and property (ii) in Proposition 2.3, we have that w is a $\mathcal{C}^1$ classical solution of (1.1) in $(0, T] \times \mathbb{R}^d$ and

$$D_x w(T, x) \subseteq \overline{B}_d\left(p_r, \frac{r}{2}\right) \subseteq \overline{B}_d(0, M) \qquad \text{for all } x \in \mathbb{R}^d.$$

Moreover, from the translation property of S_t in (2.10), it holds that

$$w(T, 0) = S_T(w_0)(0) = S_T\left(\varphi_- - S_T(\varphi_-)(0)\right)(0) = 0.$$

Thus, the continuous function $u : [0, T] \times \mathbb{R}^d \to \mathbb{R}$, defined by

$$u(t, x) = -w(T - t, -x) \qquad \text{for all } (t, x) \in [0, T] \times \mathbb{R}^d,$$

is also a $\mathcal{C}^1$ classical solution of (1.1) in $(0, T) \times \mathbb{R}^d$ with

$$u(T, \cdot) = \mathcal{T}(\varphi)(\cdot) \quad \text{and} \quad u(0, \cdot) = -w(T, -\cdot) \in \mathcal{U}_{[0,M]}.$$

In particular, $u(t, x)$ is a viscosity solution of (1.1) in $[0, T] \times \mathbb{R}^d$, so that by the uniqueness property of the semigroup map S_t, one has

$$S_T(u_0)(\cdot) = \mathcal{T}(\varphi)(\cdot), \qquad u_0(\cdot) = -w(T, -\cdot)$$

and this yields (4.21).

2. For every $\varepsilon > 0$, we select a finite subset $A_\varepsilon \subseteq [-m, m]$ such that

$$\# A_\varepsilon \;=\; \left\lfloor \frac{2^d R^d m}{\varepsilon} \right\rfloor \quad \text{and} \quad |a_i - a_j| \;\geq\; \frac{2\varepsilon}{2^d R^d} \quad \text{for all } a_i \neq a_j \in A_\varepsilon. \tag{4.22}$$

Again from the translation property of S_t in (2.10), one has

$$S_T\left(\mathcal{U}_{[m,M]}\right) \supseteq \bigcup_{a \in A_\varepsilon} S_T\left(a + \mathcal{U}_{[0,M]}\right) \;=\; A_\varepsilon + S_T\left(\mathcal{U}_{[0,M]}\right)$$

and (4.21) implies that

$$S_T\big(\mathcal{U}_{[m,M]}\big) \supseteq A_\varepsilon + \mathcal{T}\left(\mathcal{W}_r^{p_r}\right). \tag{4.23}$$

By Corollary 2.2, for every $\varepsilon > 0$ sufficiently small, there exists a set $\mathcal{G} \subseteq \mathcal{W}_r^{p_r}$ such that

$$\#\mathcal{G} \geq 2^{\beta_1 \cdot \varepsilon^{-d}}, \qquad \beta_1 = \frac{1}{3^d 2^{d^2+4d+3}\ln 2} \cdot \left(\frac{\omega_d R^{d+1} K_r}{(d+1)}\right)^d$$

and

$$\left\| D\varphi_{\llcorner \square_R} - D\phi_{\llcorner \square_R} \right\|_{\mathbf{L}^1(\square_R)} \geq 2\varepsilon \qquad \text{for all } \varphi \neq \phi \in \mathcal{G}.$$

Since $D\mathcal{T}(\varphi)(x) = D\varphi(x)$ for all $x \in \mathbb{R}^d$,

$$\left\| D\mathcal{T}(\varphi)_{\llcorner \square_R} - D\mathcal{T}(\phi)_{\llcorner \square_R} \right\|_{\mathbf{L}^1(\square_R)} \geq 2\varepsilon \qquad \text{for all } \varphi \neq \phi \in \mathcal{G}.$$

Recalling (4.22), we have

$$\left\| f_{\llcorner \square_R} - g_{\llcorner \square_R} \right\|_{\mathbf{W}^{1,1}(\square_R)} \geq 2\varepsilon \quad \text{for all } f \neq g \in A_\varepsilon + \mathcal{T}\left(\mathcal{W}_r^{p_r}\right)$$

and thus

$$\mathcal{H}_\varepsilon\left(S_{T,R}(\mathcal{U}_{[m,M]}) \Big| \mathbf{W}^{1,1}(\square_R)\right) \geq \log_2\left(\#A_\varepsilon \cdot \#\mathcal{G}\right) = \log_2\left(\left\lfloor \frac{2^d R^d m}{\varepsilon} \right\rfloor\right) + \frac{\beta_1}{\varepsilon^d}.$$

Finally, by choosing $r = \dfrac{\varepsilon}{R^-}$ with $R^- := \dfrac{\omega_d \cdot R^d}{(d+1)2^{d+9}}$, we compute

$$K_r = \frac{\lambda_M}{4TR^-} \cdot \frac{\varepsilon}{\Phi_M\left(\frac{\varepsilon}{R^-}\right)}, \qquad \beta_1 = \frac{1}{8\ln 2} \cdot \left(\frac{8R\lambda_M}{3T}\right)^d \cdot \left(\frac{\varepsilon}{\Phi_M\left(\frac{\varepsilon}{R^-}\right)}\right)^d$$

and get

$$\mathcal{H}_\varepsilon\left(S_{T,R}(\mathcal{U}_{[m,M]}) \Big| \mathbf{W}^{1,1}(\square_R)\right) \geq \frac{1}{8\ln 2} \cdot \left(\frac{8R\lambda_M}{3T}\right)^d \cdot \left(\Phi_M\left(\frac{\varepsilon}{R^-}\right)\right)^{-d} + \log_2\left(\left\lfloor \frac{2^d R^d m}{\varepsilon} \right\rfloor\right).$$

This particularly yields the first inequality in (4.4) for every $\varepsilon > 0$ sufficiently small. □

5 A counter-example

In this section, we provide an example to show that if the Hamiltonian $H \in \mathcal{C}^2(\mathbb{R}^2)$ satisfies **(H1)** but not **(H2)** then Theorem 1.2 fails in general. Consider a smooth, coercive and strictly convex Hamiltonian

$$H(p) = \frac{3^3}{4^4} \cdot p_1^4 + p_2^2, \qquad p = (p_1, p_2) \in \mathbb{R}^2.$$

The function H does not satisfy **(H2)** as

$$\lim_{p_1\to 0}\left\langle \frac{DH(p_1,p_1^2)-DH(0,0)}{|DH(p_1,p_1^2)-DH(0,0)|},\frac{(p_1,p_1^2)}{|(p_1,p_1^2)|}\right\rangle=\lim_{p_1\to 0}\left\langle \frac{\left(\frac{3^3}{4^3}\cdot p_1^3,2p_1^2\right)}{\left|\left(\frac{3^3}{4^3}\cdot p_1^3,2p_1^2\right)\right|},\frac{(p_1,p_1^2)}{|(p_1,p_1^2)|}\right\rangle=0.$$

The associated Lagrangian L of H is computed by

$$L(q) \ = \ |q_1|^{\frac{4}{3}}+q_2^2 \qquad \text{for all } q=(q_1,q_2)\in\mathbb{R}^2.$$

For any given $\bar{q}=(\bar{q}_1,\bar{q}_2)\in\mathbb{R}^2$, one has

$$\begin{aligned} L(q) \ = \ L(q-\bar{q}) \qquad &\Longleftrightarrow \qquad |q_1|^{4/3}+q_2^2 \ = \ |q_1-\bar{q}_1|^{4/3}+(q_2-\bar{q}_2)^2 \\ &\Longleftrightarrow \qquad q_2 \ = \ \frac{|q_1-\bar{q}_1|^{4/3}-|q_1|^{4/3}}{2\bar{q}_2}+\frac{\bar{q}_2}{2}. \end{aligned}$$

Let $\gamma_{\bar{q}}:\mathbb{R}\to\mathbb{R}$ be such that

$$\gamma_{\bar{q}}(s) \ = \ \frac{|s-\bar{q}_1|^{4/3}-|s|^{4/3}}{2\bar{q}_2}+\frac{\bar{q}_2}{2} \qquad \text{for all } s\in\mathbb{R}.$$

In particular, assume that $\bar{q}_2=|\bar{q}_1|^{2/3}$ with $|\bar{q}_1|=\delta$ for some $\delta>0$, it holds

$$\gamma_{\bar{q}}(0) \ = \ \bar{q}_2, \qquad \gamma_{\bar{q}}(\bar{q}_1) \ = \ 0,$$

and the following curve which connects two points $(0,\bar{q}_2)$ and $(\bar{q}_1,0)$

$$\Gamma_{\bar{q}} \ = \ \begin{cases} \{(s,\gamma_{\bar{q}}(s)):s\in[0,\delta]\} \ \subset \ [0,\delta]\times[0,\delta^{2/3}] & \text{if } \ \bar{q}_1=\delta>0 \\ \\ \{(s,\gamma_{\bar{q}}(s)):s\in[-\delta,0]\} \ \subset \ [-\delta,0]\times[0,\delta^{2/3}] & \text{if } \ \bar{q}_1=-\delta<0 \end{cases}$$

has a length $>\delta^{2/3}$. From this observation, we shall construct an initial datum $u_0\in\mathbf{Lip}(\mathbb{R}^2)$ such that both $D_xu(1,\cdot)$ and $\mathbf{b}(1,\cdot)=DH(D_xu(1,\cdot))$ do not have locally bounded variation where u is the solution of (1.1) with $u(0,\cdot)=u_0$.

Step 1: For given $0<\ell<1$, we first construct an initial datum $\bar{u}\in\mathbf{Lip}(\mathbb{R}^2)$ with $\mathrm{Lip}[\bar{u}]\leq 1$ such that

$$\operatorname{supp}(\bar{u})\subset[-2\ell,2\ell], \qquad |\mathbf{b}(1,\cdot)|([-\ell,\ell]^2),|D_xu(1,\cdot)|([-\ell,\ell]^2)\geq 1 \tag{5.1}$$

where u is the solution of (1.1) with $u(0,\cdot)=\bar{u}$. For every $0<\delta<\ell$, we consider the periodic lattice

$$y_\iota=\left(\iota_1\delta,\iota_2\delta^{2/3}\right), \qquad \iota\in\mathcal{Z}_2:=\left\{(\iota_1',\iota_2')\in\mathbb{Z}^2:\iota_1'+\iota_2'\in 2\mathbb{Z}\right\}$$

and the corresponding regions

$$\begin{aligned} \Omega_\iota \ &= \ \{x\in\mathbb{R}^2:L(x-y_\iota)<L(x-y_{\iota'}) \quad \text{for all } \iota'\neq\iota\} \\ &= \ y_\iota+\left\{q\in\mathbb{R}^2:L(q)<L(q+y_\iota-y_{\iota'}) \quad \text{for all } \iota'\neq\iota\right\} \\ &\subseteq \ y_\iota+[-\delta,\delta]\times[-\delta^{2/3},\delta^{2/3}]. \end{aligned}$$

with $\partial\Omega_\iota = \left[y_\iota + (\Gamma_{\bar q^+} \cup \Gamma_{\bar q^-})\right] \bigcup \left[y_{\iota-(1,1)} + \Gamma_{\bar q^+}\right] \bigcup \left[y_{\iota-(-1,1)} + \Gamma_{\bar q^-}\right]$ and $\bar q^\pm = (\pm\delta, \delta^{2/3})$. Let $g_0 : \mathbb{R}^2 \to \mathbb{R}$ and $g_1 : \mathcal{Z}_2 \to \mathbb{R}$ be such that

$$\begin{cases} g_1(x) & = L(x - y_\iota), \qquad x \in \Omega_\iota, \iota \in \mathcal{Z}_2, \\ g_0(y) & = \max_{x\in\mathbb{R}^2} \{g_1(x) - L(x-y)\}, \quad y \in \mathbb{R}^2. \end{cases}$$

Notice that both g_0 and g_1 are Lipschitz with Lipschitz constant

$$M_\delta = \sup_{q\in[-\delta,\delta]\times[-\delta^{2/3},\delta^{2/3}]} |DL(q)| = O(\delta^{1/3}).$$

Thus, for $\delta > 0$ sufficiently small, one can construct $\bar u \in \mathbf{Lip}(\mathbb{R}^2)$ with $\mathrm{Lip}[\bar u] \le M_\delta$ and

$$\mathrm{supp}(\bar u) \subset [-2\ell, 2\ell], \qquad \bar u(y) = g_0(y) \text{ for all } y \in \left[-\frac{3\ell}{2}, \frac{3\ell}{2}\right]^2.$$

Let u be the solution of (1.1) with $u(0,\cdot) = \bar u$. At time $t = 1$, we have

$$u(1,x) = \min_{y\in\mathbb{R}^2} \{\bar u(y) + L(x-y)\} = \bar u(y_x) + L(x - y_x)$$

for some $y_x \in \overline{B}(x, \Lambda_{M_\delta})$ with $\Lambda_{M_\delta} = \max\{|q| : L(q) \le M_\delta \cdot |q|\} = O(\delta^{1/3})$. Thus, if $\Lambda_{M_\delta} \le \frac{\ell}{2}$ then for all $x \in [-\ell,\ell]^2 \cap \Omega_\iota$, $\iota \in \mathcal{Z}_2$,

$$\begin{aligned} u(1,x) &= \min_{y\in\left[-\frac{3\ell}{2},\frac{3\ell}{2}\right]^2} \{\bar u(y) + L(x-y)\} = \min_{y\in\left[-\frac{3\ell}{2},\frac{3\ell}{2}\right]^2} \{g_0(y) + L(x-y)\} \\ &= \min_{y\in\mathbb{R}^2} \{g_0(y) + L(x-y)\} = g_1(x) = L(x - y_\iota) \end{aligned}$$

and the slope of backward optimal rays through $(1,x)$ is

$$\mathbf{b}(1,x) = DH(D_x u(1,x)) = x - y_\iota.$$

For any two vertical adjacent $y_\iota, y_{\iota'}$ with $\Omega_\iota, \Omega_{\iota'} \subset [-\ell,\ell]^2$ and $x \in \partial\Omega_\iota \cap \partial\Omega_{\iota'}$, denoting the inner normal vector to Ω_ι by $\mathbf{n}(x)$, we compute

$$\begin{cases} D_x u(1,x) & = [DL(x-y_\iota) - DL(x-y_{\iota'})] \otimes \mathbf{n}(x) \mathcal{H}^1_{\llcorner \partial\Omega_\iota\cap\partial\Omega_{\iota'}}, \\ \\ D_x \mathbf{b}(1,x) & = (y_\iota - y_{\iota'}) \otimes \mathbf{n}(x) \mathcal{H}^1_{\llcorner \partial\Omega_\iota\cap\partial\Omega_{\iota'}}. \end{cases}$$

From the definition of Ω_ι, one can show that $\mathcal{H}^1(\partial\Omega_\iota \cap \partial\Omega_{\iota'}) \ge \delta^{2/3}$ and this implies

$$|D_x\mathbf{b}(1,\cdot)|(\Omega_\iota \cup \Omega_{\iota'}), |D_x u(1,\cdot)|(\Omega_\iota \cup \Omega_{\iota'}) \ge \delta^{2/3} \cdot \mathcal{H}^1(\partial\Omega_\iota \cap \partial\Omega_{\iota'}) \ge \delta^{4/3}.$$

Moreover, since the number of open regions $\Omega_\iota \subset [-\ell,\ell]^2$ is of the order $\frac{\ell^2}{\delta^{5/3}}$, there exists a constant $C > 0$ such that

$$|D_x\mathbf{b}(1,\cdot)|([-\ell,\ell]^2), |D_x u(1,\cdot)|([-\ell,\ell]^2) \ge C \cdot \frac{\ell^2}{\delta^{5/3}} \cdot \delta^{4/3} = C \cdot \frac{\ell^2}{\delta^{1/3}}.$$

Thus, choosing $\delta > 0$ sufficiently small, we obtain (5.1).

Step 2. Consider a sequence of disjoint squares $\square_n = c_n + [0, 2^{-n}] \times [0, 2^{-n}]$ such that $\bigcup_{n\geq 1} \square_n \subset [0,1]^2$. From the previous step, for any $n \geq 1$ one can construct a sequence of functions $u_{0,n} \in \mathbf{Lip}(\mathbb{R}^2)$ with $\mathrm{Lip}[u_{0,n}] \leq 1$ such that $\mathrm{supp}(u_{0,n}) \subset \square_n$ and the solution u_n of (1.1) with $u_n(0,\cdot) = u_{0,n}(\cdot)$ satisfies

$$|D_x u_n(1,\cdot)|\left(c_n + \frac{1}{2}\cdot(\square_n - c_n)\right),\ |DH(D_x u_n(1,\cdot))|\left(c_n + \frac{1}{2}\cdot(\square_n - c_n)\right) \ \geq\ 1$$

and

$$L(x-z) \ \geq\ \min_{y\in\square_n} \{u_{0,n}(y) + L(x-y)\}, \qquad x \in \left(c_n + \frac{1}{2}\cdot(\square_n - c_n)\right), z \in \mathbb{R}^2\backslash\square_n .$$

Finally, set $u_0 = \sum_{n=1}^{\infty} u_{0,n} \in \mathbf{Lip}(\mathbb{R}^2)$. The solution u of (1.1) with $u(0,\cdot) = u_0(\cdot)$ satisfies

$$u(1,x) \ =\ u_n(1,x), \qquad x \in \left(c_n + \frac{1}{2}\cdot(\square_n - c_n)\right)$$

and this implies

$$|D_x \mathbf{b}(1,\cdot)|([0,1]^2) \ \geq\ \sum_{n=1}^{\infty} |DH(D_x u_n(1,\cdot))|(\square_n) \ \geq\ \sum_{n=1}^{\infty} 1 \ =\ +\infty.$$

Similarly, one has that $|D_x u(1,\cdot)|([0,1]^2) = +\infty$. □

Acknowledgments. This research by K.T. Nguyen was partially supported by a grant from the Simons Foundation/SFARI (521811, NTK) and by NSF grant DMS 2154201.

References

[1] G. Acosta and R. C. Dúran, An optimal Poincaré inequality in $\mathbf{L}^1$ for convex domains, *Proc. Amer. Math. Soc. Vol* **132** (2003), no.1, p. 195-202.

[2] L. Ambrosio, N. Fusco and D. Pallara, Functions of Bounded Variation and Free Discontinuity Problems, *Oxford Science Publications*, Clarendon Press, Oxford, UK, (2000).

[3] F. Ancona, P. Cannarsa and K. T. Nguyen, Quantitative compactness estimates for Hamilton-Jacobi equations, *Arch. Rat. Mech. Anal.*, **219** (2016), no. 2, 793–828.

[4] F. Ancona, P. Cannarsa and K. T. Nguyen, The compactness estimates for Hamilton Jacobi Equations depending on space, *Bull. Inst. Math. Acad. Sin. (N.S.)* **11** (2016), no. 1, 63–113.

[5] F. Ancona, O. Glass and K. T. Nguyen, Lower compactness estimates for scalar balance laws, *Comm. Pure Appl. Math* **65** (2012), no. 9, 1303–1329.

[6] F. Ancona, O. Glass and K. T. Nguyen, On quantitative compactness estimates for hyperbolic conservation laws, *Proceedings of the 14th International Conference on Hyperbolic Problems (HYP2012)*, AIMS, Springfield, MO, (2014).

[7] F. Ancona, O. Glass and K. T. Nguyen, On lower compactness estimates for general nonlinear hyperbolic systems, *Ann. Inst. H. Poincare Anal. Non Lineaire*, **32** (2015), no. 6, 1229–1257.

[8] F. Ancona, O. Glass and K. T. Nguyen, On Kolmogorov entropy compactness estimates for scalar conservation laws without uniform convexity, *SIAM J. Math. Anal.*, **51** (2019), no. 4, 3020–3051.

[9] P. Cannarsa and C. Sinestrari, Semiconcave functions, Hamilton-Jacobi equations, and optimal control, *Progress in Nonlinear Differential Equations and their Applications*, 58. Birkhäuser Boston, (2004).

[10] R. Capuani, P. Dutta and K. T. Nguyen, Metric entropy for functions of bounded total generalized variation, *SIAM J. Math. Anal.* **53** (2021), no. 1, 1168–1190.

[11] M.G. Crandall and P.-L. Lions, Viscosity solutions of Hamilton-Jacobi equations, *Trans. Amer. Math. Soc.* **277** (1983), no. 1, 1–42.

[12] C. De Lellis and F. Golse, A Quantitative Compactness Estimate for Scalar Conservation Laws, *Comm. Pure Appl. Math.* **58** (2005), no. 7, 989–998.

[13] P. Dutta and K. T. Nguyen, Covering numbers for bounded variation functions, *J. Math. Anal. Appl.* **468** (2018), no. 2, 1131–1143.

[14] A.N. Kolmogorov and V.M Tikhomirov, ε-Entropy and ε-capacity of sets in functional spaces, *Uspekhi Mat. Nauk* **14** (1959), 3-86.

[15] P.D. Lax, Course on hyperbolic systems of conservation laws, *XXVII Scuola Estiva di Fis. Mat.*, Ravello, 2002.