

# RATE OF CONVERGENCE IN PERIODIC HOMOGENIZATION OF HAMILTON-JACOBI EQUATIONS: THE CONVEX SETTING

HIROYOSHI MITAKE, HUNG V. TRAN, YIFENG YU

ABSTRACT. We study the rate of convergence of  $u^\varepsilon$ , as  $\varepsilon \rightarrow 0+$ , to  $u$  in periodic homogenization of Hamilton-Jacobi equations. Here,  $u^\varepsilon$  and  $u$  are viscosity solutions to the oscillatory Hamilton-Jacobi equation and its effective equation

$$(C)_\varepsilon \quad \begin{cases} u_t^\varepsilon + H\left(\frac{x}{\varepsilon}, Du^\varepsilon\right) = 0 & \text{in } \mathbb{R}^n \times (0, \infty), \\ u^\varepsilon(x, 0) = g(x) & \text{on } \mathbb{R}^n, \end{cases}$$

and

$$(C) \quad \begin{cases} u_t + \bar{H}(Du) = 0 & \text{in } \mathbb{R}^n \times (0, \infty), \\ u(x, 0) = g(x) & \text{on } \mathbb{R}^n, \end{cases}$$

respectively. We assume that the Hamiltonian  $H = H(y, p)$  is coercive and convex in the  $p$  variable and is  $\mathbb{Z}^n$ -periodic in the  $y$  variable, and the initial data  $g$  is bounded and Lipschitz continuous. Here,  $\bar{H}$  is the effective Hamiltonian.

We prove that

(i)

$$u^\varepsilon(x, t) \geq u(x, t) - C\varepsilon \quad \text{for all } (x, t) \in \mathbb{R}^n \times [0, \infty),$$

where  $C$  depends only on  $H$  and  $\|Dg\|_{L^\infty(\mathbb{R}^n)}$ ;

(ii) For fixed  $(x, t) \in \mathbb{R}^n \times (0, \infty)$ , if  $u$  is differentiable at  $(x, t)$  and  $\bar{H}$  is twice differentiable at  $p = Du(x, t)$ , then

$$u^\varepsilon(x, t) \leq u(x, t) + \tilde{C}_p t \varepsilon + C\varepsilon$$

provided that  $g \in C^2(\mathbb{R}^n)$  with  $\|g\|_{C^2(\mathbb{R}^n)} < \infty$ . The constant  $\tilde{C}_p$  depends only on  $H, \bar{H}, p$  and  $g$ . If  $g$  is only Lipschitz continuous, then the above inequality in (ii) is changed into  $u^\varepsilon(x, t) \leq u(x, t) + C_p \sqrt{t\varepsilon} + C\varepsilon$ .

When  $n = 2$  and  $H$  is positively homogeneous in  $p$  of some fixed degree  $k \geq 1$ , utilizing the Aubry-Mather theory, we obtain the optimal convergence rate  $O(\varepsilon)$

$$|u^\varepsilon(x, t) - u(x, t)| \leq C\varepsilon \quad \text{for all } (x, t) \in \mathbb{R}^2 \times [0, \infty).$$

Here  $C$  depends only on  $H$  and  $\|Dg\|_{L^\infty(\mathbb{R}^2)}$ .

When  $n = 1$ , the optimal convergence rate  $O(\varepsilon)$  is established for any coercive and convex  $H$ .

The convergence rate turns out to have deep connections with the dynamics of the underlying Hamiltonian system and the shape of the effective Hamiltonian  $\bar{H}$ . Some related results and counter-examples are obtained as well.

---

2010 *Mathematics Subject Classification.* 35B10 35B27 35B40, 35F21 49L25 .

*Key words and phrases.* Cell problems; periodic homogenization; optimal rate of convergence; first-order Hamilton-Jacobi equations; viscosity solutions.

The work of HM was partially supported by the JSPS grants: KAKENHI #15K17574, #26287024, #16H03948. The work of HT is partially supported by NSF grant DMS-1664424. The work of YY is partially supported by NSF CAREER award #1151919.

## 1. INTRODUCTION

We first give a brief description of the periodic homogenization theory for Hamilton-Jacobi equations. For each  $\varepsilon > 0$ , let  $u^\varepsilon \in C(\mathbb{R}^n \times [0, \infty))$  be the viscosity solution to  $(C)_\varepsilon$ . Here, the Hamiltonian  $H = H(y, p) : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$  is a given continuous function, which satisfies

- (H1) For each  $p \in \mathbb{R}^n$ ,  $y \mapsto H(y, p)$  is  $\mathbb{Z}^n$ -periodic.
- (H2)  $H$  is coercive in  $p$ , that is, uniformly for  $y \in \mathbb{T}^n := \mathbb{R}^n/\mathbb{Z}^n$ ,

$$\lim_{|p| \rightarrow \infty} H(y, p) = +\infty.$$

The initial data  $g \in \text{BUC}(\mathbb{R}^n)$ , the set of bounded, uniformly continuous functions on  $\mathbb{R}^n$ .

It was proved in [16] that, under assumptions (H1)–(H2),  $u^\varepsilon$  converges to  $u$  locally uniformly on  $\mathbb{R}^n \times [0, \infty)$  as  $\varepsilon \rightarrow 0$ , and  $u$  solves the effective equation (C). The effective Hamiltonian  $\bar{H} \in C(\mathbb{R}^n)$  is determined in a nonlinear way by  $H$  through the cell problem as following. For each  $p \in \mathbb{R}^n$ , it was shown in [16] that there exists a unique constant  $\bar{H}(p) \in \mathbb{R}$  such that the following cell (ergodic) problem has a continuous viscosity solution

$$(E)_p \quad H(y, p + Dv) = \bar{H}(p) \quad \text{in } \mathbb{T}^n.$$

If needed, we write  $v = v(y, p)$  or  $v = v_p(y)$  to demonstrate clearly the nonlinear dependence of  $v$  on  $p$ . It is worth mentioning that in general  $v(y, p)$  is not unique even up to additive constants.

Our main goal in this paper is to obtain rate of convergence of  $u^\varepsilon$  to  $u$ , that is, an optimal bound for  $\|u^\varepsilon - u\|_{L^\infty(\mathbb{R}^n \times [0, T])}$  for any given  $T > 0$  as  $\varepsilon \rightarrow 0+$ . Heuristically, owing to the two-scale asymptotic expansion,

$$u^\varepsilon(x, t) \approx u(x, t) + \varepsilon v\left(\frac{x}{\varepsilon}, Du(x, t)\right) + O(\varepsilon^2), \quad (1.1)$$

and hence, convergence rate looks like  $O(\varepsilon)$  which, if achievable, is clearly optimal. However, it is basically hopeless to justify this expansion due to two reasons:

- (1) in general, there does not exist a continuous selection of  $v(y, p)$  (or even supersolutions) with respect to  $p$ , let alone Lipschitz continuous selection;
- (2) the solution  $u(x, t)$  to (C) is only Lipschitz in  $(x, t)$ , and is usually not  $C^1$ , let alone  $C^2$ .

Up to now, the best known convergence rate is  $O(\varepsilon^{1/3})$  obtained in [9] through a kind of weak formulation of (1.1) using the ergodic problems and their discounted approximations. To be more precise, [9] deals with stationary problems, but the approach can be easily adjusted to handle the Cauchy problem. Point (1) above poses an essential obstacle to improve the convergence rate by modifying the method in [9]. See Section 5 for details. See also [1, 7, 17, 18] and the references therein for related results on rate of convergence of first order Hamilton-Jacobi equations in various settings.

**1.1. Main results.** We now describe our main results. In this paper, we assume further that

(H3) For each  $y \in \mathbb{T}^n$ ,  $p \mapsto H(y, p)$  is convex.

*Multi dimensions.* For  $n \geq 1$ , denote

$$S = \{p \in \mathbb{R}^n : \overline{H} \text{ is twice differentiable at } p\}.$$

**Theorem 1.1.** *Assume (H1)–(H3) and  $g \in \text{BUC}(\mathbb{R}^n) \cap \text{Lip}(\mathbb{R}^n)$ . For  $\varepsilon > 0$ , let  $u^\varepsilon$  be the viscosity solution to (C) $_\varepsilon$ . Let  $u$  be the viscosity solution to (C). Then,*

(i)

$$u^\varepsilon(x, t) \geq u(x, t) - C\varepsilon \quad \text{for all } (x, t) \in \mathbb{R}^n \times [0, \infty). \quad (1.2)$$

The constant  $C > 0$  in (i) and (ii) below depends only on  $H$  and  $\|Dg\|_{L^\infty(\mathbb{R}^n)}$ .

(ii) For fixed  $(x, t) \in \mathbb{R}^n \times (0, \infty)$ , if  $u$  is differentiable at  $(x, t)$  and  $p = Du(x, t) \in S$ , then

$$u^\varepsilon(x, t) \leq u(x, t) + C_p \sqrt{t\varepsilon} + C\varepsilon. \quad (1.3)$$

Here  $C_p > 0$  is a constant depending on  $H, \overline{H}, p$  and  $\|Dg\|_{L^\infty(\mathbb{R}^n)}$ . If we further assume that the initial data  $g \in C^2(\mathbb{R}^n)$  with  $\|g\|_{C^2(\mathbb{R}^n)} < \infty$ , then

$$u^\varepsilon(x, t) \leq u(x, t) + \tilde{C}_p t\varepsilon + C\varepsilon. \quad (1.4)$$

Here  $\tilde{C}_p$  is a constant depending on  $H, \overline{H}, p$  and  $\|g\|_{C^2(\mathbb{R}^n)}$ .

**Remark 1.** In general, it is probably impossible to tell whether  $\overline{H}$  is twice differentiable at a given  $p$ . Nevertheless, under assumptions (H1)–(H3), we have that  $\overline{H}$  is convex and coercive. The Alexandrov theorem yields that  $\mathbb{R}^n \setminus S$  has zero Lebesgue measure. Thus, Theorem 1.1 sort of implies that for “typical”  $g \in C^2$  and Hamiltonian  $H$ , the convergence rate is  $O(\varepsilon)$  for “most”  $(x, t) \in \mathbb{R}^n \times (0, T]$  for fixed  $T > 0$ , which is optimal. Besides the shape of  $\overline{H}$ , the upper bound of  $u^\varepsilon - u$  is also closely related to a challenging issue in weak KAM theory (or Aubry-Mather theory), which is to identify how fast the average slope of a backward characteristic defined in Definition 1 (or orbits in the Mather set)

$$\frac{\xi(t)}{t}$$

tends to  $D\overline{H}(p)$  as  $t \rightarrow -\infty$  for  $p \in \mathbb{R}^n$  and  $n > 1$ . See Lemma 2.1 for details. When  $n \geq 3$ , except for some special cases, the above statistical type upper bound (1.3) and (1.4) might be the best we can obtain since very little is known about the shape of  $\overline{H}$  and the structure of Aubry and Mather sets. See Remark 5 for more comments. The classical Hedlund example [4, 15] might be the only well-understood and interesting three dimensional example, where the convergence rate turns out to be  $O(\varepsilon)$  (Theorem 2.6).

It is reasonable to speculate that when  $n \geq 3$ , in general, the optimal convergence rate  $O(\varepsilon)$  cannot be achieved for every point  $(x, t)$ . However, we have not been able to find an explicit example with a fractional convergence rate, which basically involves determining the long time behavior of a trajectory in a chaotic Hamiltonian system.

**Remark 2.** If the initial data  $g \in \text{Lip}(\mathbb{R}^n)$  is concave, it is quite easy to get the optimal convergence rate  $O(\varepsilon)$  by using a simple PDE approach based on convexity of  $H$ . See Theorem 2.5.

*Two dimensions.* In two dimensions ( $n = 2$ ), orbits in the Mather set can be identified with circle homeomorphisms (see Appendix), which provide nice control of the convergence of the above average slope. Employing this special structure and adding a positive homogeneity assumption on  $H$ , we are able to establish the optimal rate of convergence.

**Theorem 1.2.** *Assume  $n = 2$ , (H1)–(H3) and  $g \in \text{BUC}(\mathbb{R}^2) \cap \text{Lip}(\mathbb{R}^2)$ . Assume further that  $H$  is positively homogeneous of degree  $k$  in  $p$  for some  $k \geq 1$ , that is,  $H(y, \lambda p) = \lambda^k H(y, p)$  for all  $(\lambda, y, p) \in [0, \infty) \times \mathbb{T}^2 \times \mathbb{R}^2$ . Then,*

$$|u^\varepsilon(x, t) - u(x, t)| \leq C\varepsilon \quad \text{for all } (x, t) \in \mathbb{R}^2 \times [0, \infty). \quad (1.5)$$

Here  $C > 0$  is a constant depending only on  $H$  and  $\|Dg\|_{L^\infty(\mathbb{R}^2)}$ .

Note that  $k = 1$  corresponds to Hamiltonians associated with the front propagation, which is probably one of the most physically relevant situations in the homogenization theory. See Remark 6 for comments about other types of Hamiltonians, e.g.,  $H(y, p) = \frac{1}{2}|p|^2 + V(y)$ .

*One dimension.* In one dimensional setting, we also obtain optimal convergence rate  $O(\varepsilon)$  for general convex and coercive  $H$  through the following theorem.

**Theorem 1.3.** *Assume that  $n = 1$  and (H1)–(H3) hold. Assume further that  $g \in \text{Lip}(\mathbb{R}) \cap \text{BUC}(\mathbb{R})$ . Then,*

$$\|u^\varepsilon - u\|_{L^\infty(\mathbb{R} \times [0, \infty))} \leq C\varepsilon. \quad (1.6)$$

Here  $C$  is a constant depending only on  $H$  and  $\|g'\|_{L^\infty(\mathbb{R})}$ .

It remains a very interesting problem to determine the optimal convergence rate in the nonconvex setting when  $n = 1$ . We conjecture that the optimal convergence rate is  $\sqrt{\varepsilon}$  for general coercive  $H$ .

We would like to emphasize that the constant  $C$  in our results does not depend on the smoothness of  $H$ , which will be made clear in proofs.

*Lack of continuous selection of  $v(y, p)$  with respect to  $p$ .* This fact should be known to experts although we could not find any explicit example in the literature. In Section 5, we give an example in three dimensions where there is no continuous selection of  $v(y, p)$  with respect to  $p$ . See Theorem 5.4. This example resembles some features of the more sophisticated Hedlund example [15] (see section 2.4).

**Organization of the paper.** The paper is organized as follows. We first provide settings and simplifications that can be used throughout the paper at the beginning of Section 2. The latter part of Section 2 is devoted to the proof of Theorem 1.1. We give two further situations that upper bound (1.3) can be improved in Corollary 2.4 and Theorem 2.5. The classical Hedlund example is analyzed at the end of this section (Theorem 2.6). The proof of Theorem 1.2 for positively homogeneous

Hamiltonians in two dimensions is shown in Section 3. In Section 4, we give the proof of Theorem 1.3. In Subsection 4.1, we give a simple example to demonstrate that  $O(\varepsilon)$  is indeed the optimal rate of convergence in one dimension. We discuss the nonexistence of continuous selection  $v(y, p)$  with respect to  $p$  in Section 5. Our approaches combine tools from PDE, weak KAM theory and the Aubry-Mather theory. Some basic backgrounds are provided in Appendix (Section 6).

**Acknowledgment.** We are deeply thankful to Hitoshi Ishii, who provides us invaluable comments and suggestions, which help much in vastly improving the presentation of the paper. We also would like to thank Weinan E and Jinxin Xue for helpful comments and discussions.

## 2. SETTINGS, SIMPLIFICATIONS AND PROOF OF THEOREM 1.1

In this section, (H1)–(H3) are always in force. We always assume that  $g \in \text{BUC}(\mathbb{R}^n) \cap \text{Lip}(\mathbb{R}^n)$ .

**2.1. Settings and Simplifications.** By the comparison principle, it is straightforward that

$$\|Du^\varepsilon(x, t)\|_{L^\infty(\mathbb{R}^n)} \leq C_0.$$

Here  $C_0 > 0$  is a constant depending only on  $H$  and  $\|Dg\|_{L^\infty(\mathbb{R}^n)}$ . Accordingly, values of  $H(y, p)$  for  $|p| > C_0$  are irrelevant. Hence, without loss of generality, we may further assume that  $H$  grows quadratically, that is,

$$\frac{1}{2}|p|^2 - K_0 \leq H(y, p) \leq \frac{1}{2}|p|^2 + K_0 \quad \text{for all } (y, p) \in \mathbb{T}^n \times \mathbb{R}^n, \quad (2.1)$$

for some  $K_0 > 1$ . Then, we also have that

$$\frac{1}{2}|p|^2 - K_0 \leq \bar{H}(p) \leq \frac{1}{2}|p|^2 + K_0 \quad \text{for all } p \in \mathbb{R}^n. \quad (2.2)$$

We use (2.1) and (2.2) to get that, for each  $v_p \in \text{Lip}(\mathbb{T}^n)$  solving (E)<sub>p</sub>,

$$\|Dv_p\|_{L^\infty(\mathbb{T}^n)} \leq 2(|p| + K_0).$$

In particular,

$$\max_{\mathbb{T}^n} v_p - \min_{\mathbb{T}^n} v_p \leq 2\sqrt{n}(|p| + K_0). \quad (2.3)$$

Let  $L(y, q)$  and  $\bar{L}(q)$  be the Lagrangians (Legendre transforms) of the Hamiltonians  $H(y, p)$  and  $\bar{H}(p)$ , respectively. It is clear that

$$\frac{1}{2}|q|^2 - K_0 \leq L(y, q) \leq \frac{1}{2}|q|^2 + K_0 \quad \text{for all } (y, q) \in \mathbb{T}^n \times \mathbb{R}^n, \quad (2.4)$$

and

$$\frac{1}{2}|q|^2 - K_0 \leq \bar{L}(q) \leq \frac{1}{2}|q|^2 + K_0 \quad \text{for all } q \in \mathbb{R}^n.$$

For  $(x, t) \in \mathbb{R}^n \times (0, \infty)$ , the optimal control formula for the solution to (C)<sub>ε</sub> implies

$$u^\varepsilon(x, t) = \inf_{\substack{\varepsilon\eta(0)=x \\ \eta \in \text{AC}([-\varepsilon^{-1}t, 0])}} \left\{ g(\varepsilon\eta(-\varepsilon^{-1}t)) + \varepsilon \int_{-\varepsilon^{-1}t}^0 L(\eta(s), \dot{\eta}(s)) ds \right\}. \quad (2.5)$$

Here,  $AC([-\varepsilon^{-1}t, 0])$  is the set of all absolutely continuous functions on  $[-\varepsilon^{-1}t, 0]$  with values in  $\mathbb{R}^n$ .

**2.2. Proof of Theorem 1.1.** We divide the proof into two parts. We first derive the lower bound (1.2).

*Proof of lower bound (1.2).* Let us mention right away that to get (1.2), we only need  $g \in BUC(\mathbb{R}^n) \cap Lip(\mathbb{R}^n)$ . Throughout this proof,  $C, C_1$  denote constants depending only on  $\|Dg\|_{L^\infty(\mathbb{R}^n)}$  and  $K_0$ , which might change from line to line.

By scaling and translation, it suffices to prove that (1.2) holds for  $(x, t) = (0, 1)$ . In other words, we aim to show

$$u^\varepsilon(0, 1) - u(0, 1) \geq -C\varepsilon. \quad (2.6)$$

Without loss of generality, we may assume that  $g(0) = 0$  by considering  $\tilde{g} = g(x) - g(0)$ . Hence

$$|g(x)| \leq C|x| \quad \text{for all } x \in \mathbb{R}^n. \quad (2.7)$$

The optimal control formula (2.5) gives us that

$$u^\varepsilon(0, 1) = \inf_{\substack{\eta(0)=0 \\ \eta \in AC([-\varepsilon^{-1}, 0])}} \left\{ g(\varepsilon\eta(-\varepsilon^{-1})) + \varepsilon \int_{-\varepsilon^{-1}}^0 L(\eta(t), \dot{\eta}(t)) dt \right\}.$$

Due to (2.4) and Jensen's inequality,

$$\varepsilon \int_{-\varepsilon^{-1}}^0 L(\eta(t), \dot{\eta}(t)) dt \geq \varepsilon \int_{-\varepsilon^{-1}}^0 \left( \frac{|\dot{\eta}(t)|^2}{2} - K_0 \right) dt \geq \frac{1}{2}\varepsilon^2 |\eta(-\varepsilon^{-1})|^2 - K_0.$$

Owing to (2.7), there exists  $C > 0$  such that

$$u^\varepsilon(0, 1) = \inf_{\substack{\eta(0)=0, \\ \varepsilon|\eta(-\varepsilon^{-1})| \leq C}} \left\{ g(\varepsilon\eta(-\varepsilon^{-1})) + \varepsilon \int_{-\varepsilon^{-1}}^0 L(\eta(t), \dot{\eta}(t)) dt \right\}. \quad (2.8)$$

Clearly, there exists  $C_1 > 0$  such that for any  $|q| \leq C$ ,

$$\bar{L}(q) = \sup_{p \in \mathbb{R}^n} \{p \cdot q - \bar{H}(p)\} = \sup_{|p| \leq C_1} \{p \cdot q - \bar{H}(p)\}. \quad (2.9)$$

For  $p \in \mathbb{R}^n$ , let  $v_p \in Lip(\mathbb{T}^n)$  be a viscosity solution to  $(E)_p$  such that  $v_p(0) = 0$ . Then for any Lipschitz continuous curve  $\eta : [-\varepsilon^{-1}, 0] \rightarrow \mathbb{R}^n$ ,

$$\int_{-\varepsilon^{-1}}^0 L(\eta(t), \dot{\eta}(t)) + \bar{H}(p) dt \geq p \cdot \eta(0) - p \cdot \eta(-\varepsilon^{-1}) + v_p(\eta(0)) - v_p(\eta(-\varepsilon^{-1})).$$

The above fact is easily seen in case  $v_p$  is smooth as

$$\begin{aligned} & \int_{-\varepsilon^{-1}}^0 L(\eta(t), \dot{\eta}(t)) + \bar{H}(p) dt = \int_{-\varepsilon^{-1}}^0 L(\eta(t), \dot{\eta}(t)) + H(\eta(t), p + Dv_p(\eta(t))) dt \\ & \geq \int_{-\varepsilon^{-1}}^0 \dot{\eta}(t) \cdot (p + Dv_p(\eta(t))) dt = p \cdot \eta(0) - p \cdot \eta(-\varepsilon^{-1}) + v_p(\eta(0)) - v_p(\eta(-\varepsilon^{-1})). \end{aligned}$$

Otherwise, it can be obtained by using standard convolutions of  $v_p$ . See [13] for example. Accordingly, for  $\eta(0) = 0$  and  $\varepsilon |\eta(-\varepsilon^{-1})| \leq C$ , in light of (2.3) and (2.9),

$$\begin{aligned} \varepsilon \int_{-\varepsilon^{-1}}^0 (L(\eta(t), \dot{\eta}(t)) dt &\geq \sup_{p \in \mathbb{R}^n} \{p \cdot (-\varepsilon \eta(-\varepsilon^{-1})) - \overline{H}(p) + \varepsilon v_p(0) - \varepsilon v_p(\eta(-\varepsilon^{-1}))\} \\ &\geq \sup_{|p| \leq C_1} \{p \cdot (-\varepsilon \eta(-\varepsilon^{-1})) - \overline{H}(p) + \varepsilon v_p(0) - \varepsilon v_p(\eta(-\varepsilon^{-1}))\} \\ &\geq \overline{L}(-\varepsilon \eta(-\varepsilon^{-1})) - C\varepsilon. \end{aligned}$$

We combine the above with (2.8) to yield

$$\begin{aligned} u^\varepsilon(0, 1) &\geq \inf_{\substack{\eta(0)=0, \\ \varepsilon |\eta(-\varepsilon^{-1})| \leq C}} \{g(\varepsilon \eta(-\varepsilon^{-1})) + \overline{L}(-\varepsilon \eta(-\varepsilon^{-1}))\} - C\varepsilon \\ &\geq \inf_{y \in \mathbb{R}^n} \{g(y) + \overline{L}(-y)\} - C\varepsilon \\ &= u(0, 1) - C\varepsilon. \end{aligned}$$

The last equality in the above holds thanks to the Hopf-Lax formula for  $u$ .  $\square$

We now proceed to prove upper bounds (1.3) and (1.4). We first give a definition on backward characteristics (see calibrated curves in [13]).

**Definition 1.** For each  $p \in \mathbb{R}^n$ , let  $v_p \in \text{Lip}(\mathbb{T}^n)$  be a viscosity solution to the cell problem  $(E)_p$ . Then,  $\xi : (-\infty, 0] \rightarrow \mathbb{R}^n$  is called a backward characteristic of  $v_p$  if

$$p \cdot \xi(t_1) + v_p(\xi(t_1)) - p \cdot \xi(t_2) - v_p(\xi(t_2)) = \int_{t_2}^{t_1} L(\xi(t), \dot{\xi}(t)) + \overline{H}(p) dt \quad (2.10)$$

for all  $t_2 < t_1 \leq 0$ .

If  $\xi$  is defined on  $\mathbb{R}$  and the above equality holds for all  $t_1, t_2 \in \mathbb{R}$  with  $t_2 < t_1$ ,  $\xi$  is called a global characteristic.

Note that for a curve  $\xi$  to be a backward characteristic (or global characteristic), it suffices to have (2.10) for  $t_1 = 0$  and a sequence of  $t_{2m} \rightarrow -\infty$  as  $m \rightarrow +\infty$  (or two sequences  $t_{1m} \rightarrow +\infty$  and  $t_{2m} \rightarrow -\infty$  as  $m \rightarrow +\infty$ ).

By approximating  $H$  with smooth and strictly convex Hamiltonians, it is easy to show that for each  $p \in \mathbb{R}^n$ , there exists a viscosity solution  $v_p \in \text{Lip}(\mathbb{T}^n)$  of  $(E)_p$  such that for every  $y \in \mathbb{R}^n$ , there is a backward characteristic  $\xi_y$  of  $v_p$  with  $\xi_y(0) = y$ . The following lemma is a key step toward proving (1.3) and (1.4).

**Lemma 2.1.** Fix  $(x, t) \in \mathbb{R}^n \times (0, \infty)$ . Assume that  $u$  is differentiable at  $(x, t)$  and  $\overline{H}$  is differentiable at  $p$  for  $p = Du(x, t)$ . Suppose that there exist a viscosity solution  $v_p \in \text{Lip}(\mathbb{T}^n)$  of  $(E)_p$  and a backward characteristic  $\xi : (-\infty, 0] \rightarrow \mathbb{R}^n$  of  $v_p$  such that, for some given  $C_p > 0$  and  $\alpha \in (0, 1]$ ,

$$\left| \frac{\xi(s) - \xi(0)}{s} - D\overline{H}(p) \right| \leq \frac{C_p}{|s|^\alpha} \quad \text{for all } s < 0.$$

Then

$$u^\varepsilon(x, t) \leq u(x, t) + CC_p t^{1-\alpha} \varepsilon^\alpha + C\varepsilon. \quad (2.11)$$

If we further assume that the initial data  $g \in C^2(\mathbb{R}^n)$  with  $M = \|D^2g\|_{C(\mathbb{R}^n)} < \infty$ , then the above bound can be improved to

$$u^\varepsilon(x, t) \leq u(x, t) + MC_p^2 t^{2(1-\alpha)} \varepsilon^{2\alpha} + C\varepsilon. \quad (2.12)$$

*Proof.* Note that  $\|Du\|_{L^\infty(\mathbb{R}^n \times [0, \infty))} \leq \|Dg\|_{L^\infty(\mathbb{R}^n)}$ . It suffices to prove the above for  $(x, t) = (0, t)$ . By the Hopf-Lax formula,

$$\begin{aligned} u(0, t) &= \min_{y \in \mathbb{R}^n} \{g(y) + t\bar{L}(-t^{-1}y)\} \\ &= g(y_0) + t\bar{L}(-t^{-1}y_0) \end{aligned}$$

for some  $y_0 \in \mathbb{R}^n$ . Then  $p = Du(0, t) \in \partial\bar{L}(-t^{-1}y_0)$ . Hence,  $-t^{-1}y_0 = D\bar{H}(p)$  and also

$$t\bar{L}(-t^{-1}y_0) = -y_0 \cdot p - t\bar{H}(p).$$

Let  $v_p$  and  $\xi$  be the viscosity solution and its backward characteristic from the assumption. By periodicity, we may assume that  $\xi(0) \in [0, 1]^n$ . By our assumption,

$$|y_0 - \varepsilon\xi(-\varepsilon^{-1}t) + \varepsilon\xi(0)| \leq C_p t^{1-\alpha} \varepsilon^\alpha,$$

and hence

$$|y_0 - \varepsilon\xi(-\varepsilon^{-1}t)| \leq C_p t^{1-\alpha} \varepsilon^\alpha + C\varepsilon.$$

We use the above and optimal control formula of  $u^\varepsilon(0, t)$  to compute that

$$\begin{aligned} u^\varepsilon(0, t) &\leq u^\varepsilon(\varepsilon\xi(0), t) + C\varepsilon \leq g(\varepsilon\xi(-\varepsilon^{-1}t)) + \varepsilon \int_{-\varepsilon^{-1}t}^0 L(\xi(s), \dot{\xi}(s)) ds + C\varepsilon \\ &= g(\varepsilon\xi(-\varepsilon^{-1}t)) - t\bar{H}(p) + p \cdot (-\varepsilon\xi(-\varepsilon^{-1}t)) + p \cdot (\varepsilon\xi(0)) \\ &\quad - \varepsilon v_p(\xi(-\varepsilon^{-1}t)) + \varepsilon v_p(\xi(0)) + C\varepsilon \\ &\leq g(y_0) + (-y_0) \cdot p - t\bar{H}(p) + CC_p t^{1-\alpha} \varepsilon^\alpha + C\varepsilon \\ &= u(0, t) + CC_p t^{1-\alpha} \varepsilon^\alpha + C\varepsilon. \end{aligned}$$

Next we prove (2.12). If  $g \in C^2(\mathbb{R}^n)$ , then  $p = Dg(y_0)$ . Accordingly, we can refine the above calculation as following

$$\begin{aligned} u^\varepsilon(0, t) &\leq g(\varepsilon\xi(-\varepsilon^{-1}t)) - t\bar{H}(p) + p \cdot (-\varepsilon\xi(-\varepsilon^{-1}t)) + p \cdot (\varepsilon\xi(0)) \\ &\quad - \varepsilon v_p(\xi(-\varepsilon^{-1}t)) + \varepsilon v_p(\xi(0)) + C\varepsilon \\ &\leq g(\varepsilon\xi(-\varepsilon^{-1}t)) + Dg(y_0) \cdot (-\varepsilon\xi(-\varepsilon^{-1}t)) - t\bar{H}(p) + C\varepsilon \\ &\leq g(y_0) + Dg(y_0) \cdot (-y_0) + \frac{M}{2} |y_0 - \varepsilon\xi(-\varepsilon^{-1}t)|^2 - t\bar{H}(p) + C\varepsilon \\ &\leq g(y_0) + p \cdot (-y_0) - t\bar{H}(p) + MC_p^2 t^{2(1-\alpha)} \varepsilon^{2\alpha} + C\varepsilon \\ &= u(0, t) + MC_p^2 t^{2(1-\alpha)} \varepsilon^{2\alpha} + C\varepsilon. \end{aligned}$$

□



**Remark 3.** From the above argument, it is easy to see that to obtain (2.12), the  $C^2$  assumption can be relaxed to semi-concave assumption on  $g$ , i.e.,  $g(x) - C|x|^2$  is concave for some  $C > 0$ .

Moreover, if  $\bar{H}$  is not differentiable at  $p = Du(x, t)$ , then  $q_0 = -t^{-1}y_0 \in \partial\bar{H}(p)$ . For each  $\varepsilon > 0$ , we may replace the backward characteristic by an ‘‘approximate characteristic’’  $\xi_\varepsilon : [-\frac{t}{\varepsilon}, 0] \rightarrow \mathbb{R}^n$  satisfying that

$$\int_{-\frac{t}{\varepsilon}}^0 L(\xi_\varepsilon, \dot{\xi}_\varepsilon) + \bar{H}(p) ds \leq p \cdot \left( \xi_\varepsilon(0) - \xi_\varepsilon \left( -\frac{t}{\varepsilon} \right) \right) + C_p,$$

and

$$\left| \frac{\xi_\varepsilon(s) - \xi_\varepsilon(0)}{s} - q_0 \right| \leq \frac{C_p}{|s|^\alpha} \quad \text{for } s = -\frac{t}{\varepsilon}.$$

Sometimes, an ‘‘approximate characteristic’’ can be constructed through proper combination of several backward characteristics associated with the same  $p$ . See Case 2 in the proof of Theorem 2.6 for the Hedlund example.

Since  $\|Du\|_{L^\infty(\mathbb{R}^n \times [0, \infty))} = \|Dg\|_{L^\infty(\mathbb{R}^n)}$  and  $u$  is differentiable a.e. in  $\mathbb{R}^n \times (0, \infty)$ , by Lemma 2.1 and approximations, we have that

**Corollary 2.2.** *Assume that  $\bar{H} \in C^1(\mathbb{R}^n)$ . Assume further that for every  $|p| \leq \|Dg\|_{L^\infty(\mathbb{R}^n)}$ , there exist a viscosity solution  $v_p \in \text{Lip}(\mathbb{T}^n)$  of  $(E)_p$  and a backward characteristic  $\xi : (-\infty, 0] \rightarrow \mathbb{R}^n$  of  $v_p$  such that, for some  $C > 0$  independent of  $p$ ,*

$$\left| \frac{\xi(s) - \xi(0)}{s} - D\bar{H}(p) \right| \leq \frac{C}{|s|} \quad \text{for all } s < 0.$$

Then

$$u^\varepsilon(x, t) \leq u(x, t) + C\varepsilon \quad \text{for all } (x, t) \in \mathbb{R}^n \times [0, \infty). \quad (2.13)$$

It is well-known that, for a given backward characteristic  $\xi$  of  $v_p$ ,

$$\frac{\xi(s) - \xi(0)}{s} \rightarrow q \in \partial\bar{H}(p) \quad \text{as } s \rightarrow -\infty, \quad (2.14)$$

by passing to a subsequence if necessary. See [10, 12, 13, 14]. However, a convergence rate of (2.14) is in general not available without extra assumptions. The following lemma says that if  $\bar{H}$  is twice differentiable at  $p$ , then the exponent  $\alpha$  in the above lemma is at least  $\frac{1}{2}$ . See [14, Theorem 1] for a similar result. The proof is a simple refinement of that of (2.14).

**Lemma 2.3.** *Assume  $\bar{H}$  is twice differentiable at  $p \in \mathbb{R}^n$ , that is,  $p \in S$ . Let  $\xi : (-\infty, 0] \rightarrow \mathbb{R}^n$  be a backward characteristic of  $v_p \in \text{Lip}(\mathbb{T}^n)$ , a solution of  $(E)_p$ . Then, there exists  $C_p > 0$  depending on  $H, \bar{H}, p$  and  $\|Dg\|_{L^\infty(\mathbb{R}^n)}$  such that*

$$\left| \frac{\xi(t) - \xi(0)}{t} - D\bar{H}(p) \right| \leq \frac{C_p}{\sqrt{|t|}} \quad \text{for } t < 0.$$

*Proof.* It suffices to consider  $t \leq -1$ . Set

$$w = \frac{\xi(t) - \xi(0)}{t} - D\bar{H}(p).$$

There is nothing to prove if  $w = 0$ . We therefore may assume that  $w \neq 0$ .

By the definition of backward characteristics,

$$p \cdot \xi(0) - p \cdot \xi(t) + v_p(\xi(0)) - v_p(\xi(t)) = \int_t^0 L(\xi(s), \dot{\xi}(s)) + \overline{H}(p) ds.$$

For  $\tilde{p} \in \mathbb{R}^n$ , let  $v_{\tilde{p}} \in \text{Lip}(\mathbb{T}^n)$  be a solution to  $(E)_{\tilde{p}}$ . It is clear that

$$\tilde{p} \cdot \xi(0) - \tilde{p} \cdot \xi(t) + v_{\tilde{p}}(\xi(0)) - v_{\tilde{p}}(\xi(t)) \leq \int_t^0 L(\xi(s), \dot{\xi}(s)) + \overline{H}(\tilde{p}) ds.$$

Accordingly, for  $|\tilde{p} - p| \leq 1$ ,

$$\overline{H}(\tilde{p}) - \overline{H}(p) \geq (\tilde{p} - p) \cdot \frac{\xi(t) - \xi(0)}{t} - \frac{C(1 + |p|)}{|t|}. \quad (2.15)$$

Since  $H$  is twice differentiable at  $p$ , there exists a constant  $C_p > 0$  such that, for  $|\tilde{p} - p| \leq 1$ ,

$$\overline{H}(\tilde{p}) \leq \overline{H}(p) + D\overline{H}(p) \cdot (\tilde{p} - p) + C_p |\tilde{p} - p|^2.$$

Combine this with the above inequality to deduce that

$$C_p |\tilde{p} - p|^2 \geq (\tilde{p} - p) \cdot \left( \frac{\xi(t) - \xi(0)}{t} - D\overline{H}(p) \right) - \frac{C(1 + |p|)}{|t|}.$$

Choose  $\tilde{p} = p + \frac{1}{\sqrt{|t|}} \frac{w}{|w|}$  to complete the proof.  $\square$

We are now ready to obtain (1.3) and (1.4).

*Proof of upper bounds (1.3) and (1.4).* Inequalities (1.3) and (1.4) follow immediately from Lemmas 2.1 and 2.3.  $\square$

**2.3. Improved upper bound.** We now give some further situations that the upper bound of  $u^\varepsilon - u$  can be improved to  $O(\varepsilon)$  fully. The first situation concerns the flat part of  $\overline{H}$ .

**Corollary 2.4.** *Fix  $(x, t) \in \mathbb{R}^n \times (0, \infty)$ . Assume that  $u$  is differentiable at  $(x, t)$  with  $p_0 = Du(x, t)$ . Assume that  $p_0$  is an interior point of  $F = \{\overline{H} = \min \overline{H}\}$ , the flat part of  $\overline{H}$ .*

(i) *If we have further that*

$$\max_{y \in \mathbb{T}^n} \min_{p \in \mathbb{R}^n} H(y, p) = \min_{p \in \mathbb{R}^n} \max_{y \in \mathbb{T}^n} H(y, p), \quad (2.16)$$

then

$$|u^\varepsilon(x, t) - u(x, t)| \leq C\varepsilon.$$

Here and below,  $C$  depends only on  $H, \|Dg\|_{L^\infty(\mathbb{R}^n)}$ .

(ii) *If we do not have (2.16), then we still have*

$$|u^\varepsilon(x, t) - u(x, t)| \leq \frac{C\varepsilon}{\delta},$$

where  $\delta = \text{dist}(p_0, \partial F) > 0$ .

*Proof.* We only need to show that  $u^\varepsilon(x, t) \leq u(x, t) + C\varepsilon$ .

Let us first prove (i). From the cell problem  $(E)_p$  for  $p \in \mathbb{R}^n$ , it is easy to see that

$$\max_{y \in \mathbb{T}^n} \min_{p \in \mathbb{R}^n} H(y, p) \leq \min_{p \in \mathbb{R}^n} \bar{H}(p) \leq \min_{p \in \mathbb{R}^n} \max_{y \in \mathbb{T}^n} H(y, p).$$

Also, since  $\min_{p \in \mathbb{R}^n} H(y, p) = -L(y, 0)$  for each  $y \in \mathbb{T}^n$ , it is clear that

$$\max_{y \in \mathbb{T}^n} \min_{p \in \mathbb{R}^n} H(y, p) = \max_{y \in \mathbb{T}^n} (-L(y, 0)).$$

Since  $\bar{H}(p_0) = \min_{p \in \mathbb{R}^n} \bar{H}(p)$  and (2.16) holds, we have that

$$\bar{H}(p_0) = \max_{y \in \mathbb{T}^n} \min_{p \in \mathbb{R}^n} H(y, p) = -L(x_0, 0)$$

for some  $x_0 \in [0, 1]^n$ . Note that  $D\bar{H}(p_0) = 0$ . Set  $\xi(t) = x_0$  for  $t \leq 0$ , and observe that  $\xi$  is a backward characteristic of any solution  $v_{p_0}$  of  $(E)_{p_0}$ . We then use the fact that

$$\left| \frac{\xi(t) - \xi(0)}{t} - D\bar{H}(p_0) \right| = 0,$$

and (2.11) to conclude.

We now prove (ii). For this, we repeat the proof of Lemma 2.3 and note that, for  $|\tilde{p} - p_0| \leq \delta$ ,

$$0 = \bar{H}(\tilde{p}) - \bar{H}(p_0) \geq (\tilde{p} - p_0) \cdot \left( \frac{\xi(t) - \xi(0)}{t} \right) - \frac{C(1 + |p_0|)}{|t|}.$$

Thus,

$$\left| \frac{\xi(t) - \xi(0)}{t} \right| \leq \frac{C(1 + |p_0|)}{\delta|t|} \leq \frac{C}{\delta|t|},$$

which, together with Lemma 2.1, yields the conclusion.  $\square$

**Remark 4.** In Corollary 2.4(i), we in fact only need to assume that  $p_0 \in F$  such that  $\bar{H}$  is differentiable at  $p_0$ . Of course, when  $p_0$  is an interior point of  $F$ , we automatically have that  $\bar{H}$  is differentiable at  $p_0$  and  $D\bar{H}(p_0) = 0$ .

Also, an example of  $H$  satisfying (2.16) is a separable Hamiltonian of the form  $H(y, p) = F(p) + V(y)$  for  $(y, p) \in \mathbb{R}^n \times \mathbb{R}^n$ .

If initial data  $g$  is concave, then the optimal convergence rate  $O(\varepsilon)$  can be derived in a different way. In this situation, we only assume that  $g \in \text{Lip}(\mathbb{R}^n)$  and do not require that  $g$  is bounded.

**Theorem 2.5.** *Suppose that  $g \in \text{Lip}(\mathbb{R}^n)$  is concave. Then*

$$|u^\varepsilon(x, t) - u(x, t)| \leq C\varepsilon \quad \text{for all } (x, t) \in \mathbb{R}^n \times [0, \infty).$$

*The constant  $C$  depends on  $H, \|Dg\|_{L^\infty(\mathbb{R}^n)}$ .*

We give a PDE proof here, which might be of independent interest.

*Proof.* Since  $g$  is Lipschitz continuous and concave, we write that

$$g(x) = \inf_{z \in \mathbb{R}^n} \{p_z \cdot (x - z) + g(z)\}.$$

Here  $p_z \in D^+g(z)$  for  $z \in \mathbb{R}^n$ . It is important noting that, as  $g \in \text{Lip}(\mathbb{R}^n)$ , we always have, for  $z \in \mathbb{R}^n$ ,

$$|p_z| \leq \|Dg\|_{L^\infty(\mathbb{R}^n)}.$$

Denote by

$$\tilde{u}^\varepsilon(x, t) = \inf_{z \in \mathbb{R}^n} \left\{ p_z \cdot (x - z) + g(z) + \varepsilon v\left(\frac{x}{\varepsilon}, p_z\right) - \overline{H}(p_z)t \right\}.$$

Then,  $\tilde{u}^\varepsilon$  is a viscosity solution to  $(C)_\varepsilon$  with initial data

$$\tilde{u}^\varepsilon(x, 0) = \inf_{z \in \mathbb{R}^n} \left\{ p_z \cdot (x - z) + g(z) + \varepsilon v\left(\frac{x}{\varepsilon}, p_z\right) \right\}$$

as it is the inf of a family of solutions to  $(C)_\varepsilon$  and  $H$  is convex. Here,  $v(y, p_z)$  is a viscosity to the cell problem  $(E)_{p_z}$  with  $v(0, p_z) = 0$  for  $z \in \mathbb{R}^n$ . Since

$$\|g - \tilde{u}^\varepsilon(\cdot, 0)\|_{L^\infty(\mathbb{R}^n)} \leq C\varepsilon,$$

we use the comparison principle to imply

$$|u^\varepsilon(x, t) - \tilde{u}^\varepsilon(x, t)| \leq C\varepsilon \quad \text{for all } (x, t) \in \mathbb{R}^n \times [0, \infty).$$

Finally, by convexity of  $\overline{H}$ , we note

$$u(x, t) = \inf_{z \in \mathbb{R}^n} \left\{ p_z \cdot (x - z) + g(z) - \overline{H}(p_z)t \right\}$$

to conclude the proof.  $\square$

**2.4. Hedlund's Example.** When  $n \geq 3$ , the only well-understood and interesting example is the classical Hedlund example (see [4, Section 5] for details). Let us consider the simplest case in three dimensions ( $n = 3$ )

$$H(y, p) = \frac{1}{a(y)}|p| \quad \text{for all } (y, p) \in \mathbb{R}^3 \times \mathbb{R}^3, \quad (2.17)$$

where  $a : \mathbb{R}^3 \rightarrow [\delta, 1 + \delta]$  is a smooth  $\mathbb{Z}^3$ -periodic function satisfying

- (i)  $a \geq 1$  outside  $U_\delta(\mathcal{L})$  and  $\min_{\mathbb{R}^3} a = \delta$ ;
- (ii)  $a(y) = \delta$  if and only if  $y \in \mathcal{L}$ .

Here,

$$\mathcal{L} = \bigcup_{i=1}^3 (l_i + \mathbb{Z}^3)$$

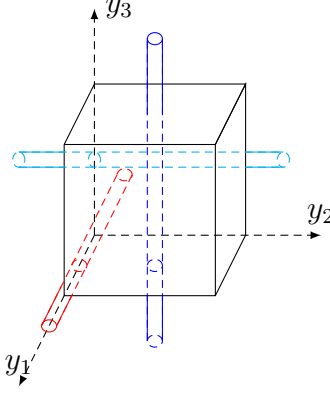
where  $l_1 = \mathbb{R} \times \{0\} \times \{0\}$ ,  $l_2 = \{0\} \times \mathbb{R} \times \{\frac{1}{2}\}$  and  $l_3 = \{\frac{1}{2}\} \times \{\frac{1}{2}\} \times \mathbb{R}$  are straight lines in  $\mathbb{R}^3$ . The constant  $\delta \in (0, 10^{-2})$  is fixed, and  $U_\delta$  is the Euclidean  $\delta$ -neighborhood of  $\mathcal{L}$ , which is basically the union of tubes. For  $1 \leq i \leq 3$ , each  $l_i$  is a minimizing geodesic for the Riemannian metric  $ds^2 = a(y)^2 \sum_{i=1}^3 dy_i^2$ . In particular, they are also trajectories in Mather sets.

It was proved in [4] (in the dual form of the stable norm) that

$$\overline{H}(p) = \frac{1}{\delta} \max \{|p_1|, |p_2|, |p_3|\}.$$

See Lemma 5.2 below for a similar formula.

Note that by the explanation at the beginning of this section,  $H$  can be treated as a superlinear Hamiltonian (that is, we can replace  $H$  by a superlinear  $\tilde{H}$  such


 FIGURE 2.1. Shape of  $U_\delta(\mathcal{L})$ 

that  $\tilde{H} \geq H$  and  $\tilde{H} = H$  when  $|p| \leq C$  for a suitable constant  $C$ ). Clearly, the corresponding Lagrangian  $L$  satisfies that

$$L(y, q) = 0 \quad \text{for } |q| \leq \frac{1}{a(y)}. \quad (2.18)$$

Define, for  $t \in \mathbb{R}$ ,

$$\xi_1(t) = \left( \frac{t}{\delta}, 0, 0 \right), \quad \xi_2(t) = \left( 0, \frac{t}{\delta}, \frac{1}{2} \right), \quad \xi_3(t) = \left( \frac{1}{2}, \frac{1}{2}, \frac{t}{\delta} \right).$$

Although the Hedlund example demonstrates the limitation of the Aubry-Mather theory in higher dimensions, we are still able to obtain the optimal convergence rate for the corresponding homogenization problem.

**Theorem 2.6.** *Let  $H$  be given by (2.17) with  $a(\cdot)$  satisfying (i)–(ii) listed right below (2.17) and  $g \in \text{BUC}(\mathbb{R}^3) \cap \text{Lip}(\mathbb{R}^3)$ . Then*

$$\|u^\varepsilon - u\|_{L^\infty(\mathbb{R}^3 \times [0, \infty))} \leq C\varepsilon.$$

Here,  $C$  is a constant depending only on  $a(\cdot)$  and  $\|Dg\|_{L^\infty(\mathbb{R}^3)}$ .

*Proof.* Fix  $(x, t) \in \mathbb{R}^3 \times (0, \infty)$ . It suffices to show that

$$u^\varepsilon(x, t) - u(x, t) \leq C\varepsilon. \quad (2.19)$$

By approximation, we may assume that  $g \in C^1$ ,  $u$  is differentiable at  $(x, t)$ . Denote  $p = Du(x, t)$ . There are two cases to be considered.

**Case 1.**  $\bar{H}$  is differentiable at  $p$ . Without loss of generality, we assume that

$$D\bar{H}(p) = \left( \frac{1}{\delta}, 0, 0 \right).$$

Then, (2.19) follows immediately from

$$\int_{-s}^0 L(\xi_1, \dot{\xi}_1) + \bar{H}(p) ds = \int_{-s}^0 \left( 0 + \frac{|p_1|}{\delta} \right) ds = p \cdot (\xi_1(0) - \xi_1(-s)) \quad \text{for all } s \geq 0,$$

and Lemma 2.1. Note that  $\xi_1$  is a global characteristic of any corrector  $v_p$  associated with  $p$  (see Definition 1 for definition of global characteristics).

**Case 2.**  $\bar{H}$  is not differentiable at  $p$ . We may assume that  $p = (1, 1, 0)$ . The proof for other  $p$  is the same. Then

$$\partial\bar{H}(p) = \left\{ \frac{1}{\delta}(\tau, 1 - \tau, 0) : \tau \in [0, 1] \right\}.$$

According to Remark 3, it suffices to construct an ‘‘approximate backward characteristic’’  $\xi_\tau : [-\frac{t}{\varepsilon}, 0] \rightarrow \mathbb{R}^3$  for a given  $q_\tau = \frac{1}{\delta}(\tau, 1 - \tau, 0) \in \partial\bar{H}(p)$  such that

$$\left| \frac{\xi_\tau(s) - \xi_\tau(0)}{s} - q_\tau \right| \leq \frac{C}{|s|} \quad \text{for } s = -\frac{t}{\varepsilon}, \quad (2.20)$$

and

$$\int_{-\frac{t}{\varepsilon}}^0 L(\xi_\tau, \dot{\xi}_\tau) + \bar{H}(p) ds \leq p \cdot \left( \xi_\tau(0) - \xi_\tau \left( -\frac{t}{\varepsilon} \right) \right) + C. \quad (2.21)$$

The idea is quite simple. We build  $\xi_\tau$  by combining a part of  $\xi_1$  with a part of a proper translation of  $\xi_2$ . Let  $k \in \mathbb{Z}$  be the integer part of  $-\frac{\tau t}{\varepsilon \delta}$ .

$$\xi_\tau(s) = \begin{cases} \left( \frac{s}{\delta}, 0, 0 \right) & \text{for } s \in [-\frac{\tau t}{\varepsilon}, 0], \\ \left( s + \frac{\tau t}{\varepsilon} \right) \left( \left( -\frac{\tau t}{\varepsilon \delta}, 0, 0 \right) - \left( k, 0, \frac{1}{2} \right) \right) + \left( -\frac{\tau t}{\varepsilon \delta}, 0, 0 \right) & \text{for } s \in [-\frac{\tau t}{\varepsilon} - 1, -\frac{\tau t}{\varepsilon}], \\ \left( k, \frac{1}{\delta} \left( s + \frac{\tau t}{\varepsilon} + 1 \right), \frac{1}{2} \right) & \text{for } s \in [-\frac{t}{\varepsilon}, -\frac{\tau t}{\varepsilon} - 1]. \end{cases}$$

It is clear from the construction that (2.20) holds. Let us now verify (2.21). Due to (2.18) and  $a = \delta$  on  $\mathcal{L}$ , we compute

$$\int_{-\frac{\tau t}{\varepsilon}}^0 L(\xi_\tau, \dot{\xi}_\tau) + \bar{H}(p) ds = \frac{\tau t}{\varepsilon \delta} = p \cdot \left( \xi_\tau(0) - \xi_\tau \left( -\frac{\tau t}{\varepsilon} \right) \right),$$

and

$$\int_{-\frac{t}{\varepsilon}}^{-\frac{\tau t}{\varepsilon} - 1} L(\xi_\tau, \dot{\xi}_\tau) + \bar{H}(p) ds = \frac{1}{\delta} \left( \frac{(1 - \tau)t}{\varepsilon} - 1 \right) = p \cdot \left( \xi_\tau \left( -\frac{\tau t}{\varepsilon} - 1 \right) - \xi_\tau \left( -\frac{t}{\varepsilon} \right) \right).$$

Then (2.21) follows from the fact that  $|\xi_\tau(-\frac{\tau t}{\varepsilon} - 1) - \xi_\tau(-\frac{\tau t}{\varepsilon})| \leq 2$  and  $|\dot{\xi}_\tau(s)| \leq 2$  for  $s \in [-\frac{\tau t}{\varepsilon} - 1, -\frac{\tau t}{\varepsilon}]$ .

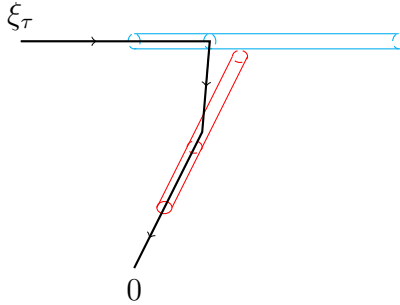


FIGURE 2.2. Approximate backward characteristic  $\xi_\tau$

□

**Remark 5.** In general, unlike the two dimensional case (see Appendix), very little is known about the structure of Aubry and Mather sets when  $n \geq 3$  due to the lack of topological restrictions. Identifying the shape of  $\overline{H}$  is extremely challenging. Again, for  $n \geq 3$ , consider the basic form of a coercive, convex and positively homogeneous Hamiltonian of degree 1 as following

$$H(y, p) = \frac{1}{a(y)}|p| \quad \text{for all } (y, p) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Here,  $a(\cdot)$  is a positive, smooth and  $\mathbb{Z}^n$ -periodic function. It is clear that the corresponding effective Hamiltonian  $\overline{H}(p)$  is also coercive, convex and positively homogeneous of degree 1. However, it is very hard to derive any further information about  $\overline{H}$ . For instance, it is not even known whether  $\overline{H}$  could be isotropic, that is,  $\overline{H}(p) = |p|$  for all  $p \in \mathbb{R}^n$ , if  $a(\cdot)$  is not constant (the answer is negative in two dimensions [5]).

By the Hopf-Lax formula, for  $(x, t) \in \mathbb{R}^n \times (0, \infty)$ ,

$$u(x, t) = \min_{y \in D_0} g(x - ty),$$

where  $D_0 = \partial \overline{H}(0)$ . It seems necessary to add certain technical assumptions about geometric properties of  $D_0$  or the underlying Hamiltonian systems in order to improve (ii) in Theorem 1.1 when  $n \geq 3$  (e.g., obtaining  $O(\varepsilon)$  for a.e.  $(x, t) \in \mathbb{R}^n \times (0, \infty)$ ). The key question is what kind of assumptions could be considered as “reasonable” and “generic” even if they are in general hard to verify. We plan to investigate this in the future.

### 3. PROOF OF THEOREM 1.2

In this section, the assumptions of Theorem 1.2 are always in force. By approximation, we may assume that, for each  $y \in \mathbb{R}^2$ ,  $p \mapsto H(y, p)$  is smooth away from the origin, and for each  $y \in \mathbb{T}^2$ , the level curve

$$\{p \in \mathbb{R}^2 : H(y, p) = 1\}$$

is strictly convex, i.e., the curvature is positive. It is worth mentioning that constant  $C$  in Theorem 1.2 does not depend on the smoothness of  $H$ . Denote

$$S_1 = \{p \in \mathbb{R}^2 : \overline{H}(p) = 1\}.$$

It is clear that  $\overline{H}$  is positively homogeneous of degree  $k$ . By [8],  $S_1$  is  $C^1$  and therefore  $\overline{H}$  is  $C^1$  away from the origin (see Theorem 6.1 in Appendix for a sketch of the proof of this interesting fact). For  $p \in S_1$ , let  $v$  be a viscosity solution to  $(E)_p$ , that is,  $v$  solves

$$H(y, p + Dv) = \overline{H}(p) = 1 \quad \text{in } \mathbb{T}^2.$$

Choose a smooth (Tonelli) Hamiltonian  $\tilde{H} : \mathbb{T}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$  such that  $\tilde{H}$  satisfies (H1)–(H3),  $\tilde{H}(y, \cdot) : \mathbb{R}^2 \rightarrow \mathbb{R}$  is strictly convex for all  $y \in \mathbb{T}^2$ , and

$$\tilde{H} = e^{H-1} \quad \text{when } H \geq \frac{1}{2}.$$

Clearly,  $\widetilde{H}(p) = e^{\overline{H}(p)-1} = 1$  and  $v$  is also a viscosity solution to

$$\widetilde{H}(y, p + Dv) = 1 \quad \text{in } \mathbb{T}^2.$$

Choose  $\xi : \mathbb{R} \rightarrow \mathbb{R}^2$  to be an orbit from the projected Mather set  $\mathcal{M}_p$  associated with  $\widetilde{H}$ . Here we lift  $\xi$  from  $\mathbb{T}^2$  to  $\mathbb{R}^2$ . Then  $v$  is  $C^1$  along  $\xi$  and for  $t \in \mathbb{R}$ ,

$$\widetilde{H}(\xi(t), p + Dv(\xi(t))) = H(\xi(t), p + Dv(\xi(t))) = 1$$

and

$$\dot{\xi}(t) = D_p \widetilde{H}(\xi(t), p + Dv(\xi(t))) = D_p H(\xi(t), p + Dv(\xi(t))). \quad (3.1)$$

Without loss of generality, we may assume that  $\xi(0) \in [0, 1]^2$ .

**Lemma 3.1.** *There exists a constant  $C > 0$  independent of  $p \in S_1$  and  $t \in \mathbb{R}$  such that*

$$|\xi(t) - \xi(0) - tD\overline{H}(p)| \leq C \quad \text{for all } t \in \mathbb{R}. \quad (3.2)$$

*Proof.* It suffices to show that

$$|\xi(t) - tD\overline{H}(p)| \leq C \quad \text{for all } |t| \geq 1.$$

Write  $D\overline{H}(p) = (a, b)$  and  $\xi(t) = (x(t), y(t))$ . Since  $p \cdot D\overline{H}(p) = k$ ,  $|D\overline{H}(p)| \geq \frac{k}{|p|}$ . Without loss of generality, we may assume that

$$a > 0 \quad \text{and} \quad \frac{b}{a} \in [0, 1].$$

Due to the identification of  $\xi$  with a circle homeomorphism (see Lemma 6.2 in Appendix) from the Aubry-Mather theory, we have that

$$\left| y(t) - \frac{b}{a}x(t) \right| \leq C.$$

We would like to remind readers that the above inequality is derived mainly based on topological arguments. In particular, the constant  $C$  does not rely on the smoothness of  $H$  (see Remark 9).

Accordingly, for  $|t| \geq 1$ ,

$$\left| \frac{\xi(t)}{t} - \frac{x(t)}{t} \left( 1, \frac{b}{a} \right) \right| \leq \frac{C}{|t|}. \quad (3.3)$$

Owing to (3.1),

$$\begin{aligned} p \cdot \xi(t) - p \cdot \xi(0) + v(\xi(t)) - v(\xi(0)) &= \int_0^t (p + Dv(\xi(s))) \cdot \dot{\xi}(s) \, ds \\ &= \int_0^t (p + Dv(\xi(s))) \cdot D_p H(\xi(s), p + Dv(\xi(s))) \, ds \\ &= kt = tp \cdot D\overline{H}(p). \end{aligned}$$

Therefore, for  $|t| \geq 1$ ,

$$\left| p \cdot \left( \frac{\xi(t)}{t} - D\overline{H}(p) \right) \right| \leq \frac{C}{|t|}. \quad (3.4)$$



Combining this with (3.3) and the fact that

$$p \cdot \left(1, \frac{b}{a}\right) = \frac{k}{a},$$

we have that

$$\left| \frac{x(t)}{t} - a \right| \leq \frac{Ca}{k|t|}.$$

Again, thanks to (3.3),

$$\begin{aligned} \left| \frac{\xi(t)}{t} - D\bar{H}(p) \right| &\leq \left| \frac{x(t)}{t} \left(1, \frac{b}{a}\right) - D\bar{H}(p) \right| + \frac{C}{|t|} \\ &= \left| \frac{x(t)}{t} - a \right| \frac{\sqrt{a^2 + b^2}}{a} + \frac{C}{|t|}. \end{aligned}$$

Therefore,

$$\left| \frac{\xi(t)}{t} - D\bar{H}(p) \right| \leq \frac{C}{|t|} \left( \frac{|D\bar{H}(p)|}{k} + 1 \right) \leq \frac{C}{|t|} \left( \frac{\max_{p' \in S_1} |D\bar{H}(p')|}{k} + 1 \right).$$

Hence (3.2) holds.  $\square$

*Proof of Theorem 1.2.* By approximation, it suffices to prove Theorem 1.2 for  $k > 1$ . Of course, we only need to prove the upper bound, that is,

$$u^\varepsilon(x, t) \leq u(x, t) + C\varepsilon \quad \text{for all } (x, t) \in \mathbb{R}^n \times [0, +\infty).$$

Thanks to Corollary 2.2, this follows immediately from the following claim.

**Claim:** For any  $p_0 \in \mathbb{R}^n$ , let  $v_0$  be a viscosity solution to (E) $_{p_0}$ . Then there exists a global characteristic  $\xi_0 : \mathbb{R} \rightarrow \mathbb{R}^2$  associated with  $v_0$  such that

$$|\xi_0(t) - \xi_0(0) - tD\bar{H}(p_0)| \leq C \quad \text{for all } t \in \mathbb{R}.$$

The constant  $C$  is the same as that in Lemma 3.1. In fact, there are two cases to be considered.

**Case 1.**  $p_0 = 0$ . Then  $D\bar{H}(p_0) = 0$ . Since  $L(y, 0) = 0$  for all  $y \in \mathbb{R}^2$ , we can simply choose  $\xi_0(s) = 0$  for  $s \in \mathbb{R}$ .

**Case 2.**  $p_0 \neq 0$ . Then  $\bar{H}(p_0) > 0$ . Choose  $\tau > 0$  satisfying that  $\tau^k = \bar{H}(p_0)^{-1}$ . Then

$$\bar{H}(\tau p_0) = 1.$$

Let  $\xi$  be a global characteristic associated with  $v = \tau v_0$  and  $p = \tau p_0$  as discussed in (3.1). Since  $D\bar{H}(\tau p) = \tau^{k-1} D\bar{H}(p)$ , we get that  $\xi_0(t) = \xi(\tau^{1-k}t)$  (equivalently,  $\xi(t) = \xi_0(\tau^{k-1}t)$ ) is a global characteristic associated with  $v_0$ . In light of Lemma 3.1, we have that

$$|\xi_0(\tau^{k-1}t) - \xi_0(0) - tD\bar{H}(\tau p_0)| \leq C \quad \text{for all } t \in \mathbb{R}.$$

By using the fact that  $D\bar{H}(\tau p_0) = \tau^{k-1} D\bar{H}(p_0)$ , we imply

$$|\xi_0(t) - \xi_0(0) - tD\bar{H}(p_0)| \leq C \quad \text{for all } t \in \mathbb{R}.$$

$\square$

**Remark 6.** The assumption of homogeneity plays a crucial role in deriving the control along a non-tangential direction in (3.4). For other important Hamiltonians, e.g., the mechanical one  $H(y, p) = \frac{1}{2}|p|^2 + V(y)$ , a reasonable question is whether, in two dimensions, upper bound (1.4) can be improved to

$$u^\varepsilon(x, t) \leq u(x, t) + C_p \sqrt{t} \varepsilon + C \varepsilon \quad \text{for a.e. } (x, t) \in \mathbb{R}^2 \times (0, +\infty).$$

Using special structures in two dimensions, this boils down to find a second derivative bound of  $\bar{H}$  along at least one non-tangential direction. More precisely, given  $p \in \mathbb{R}^2$ , if  $\bar{H}(p) > \min \bar{H}$  and  $D\bar{H}(p)$  is an irrational vector, does there exist a unit vector  $r$  such that  $r \cdot D\bar{H}(p) \neq 0$  and the function

$$w(s) = \bar{H}(p + sr)$$

satisfies that  $w(s) \leq w(0) + w'(0)s + Cs^2$ ? We will investigate this in the future.

#### 4. PROOF OF THEOREM 1.3

We are now in the one dimensional setting.

*Proof of Theorem 1.3.* Since the constant  $C$  does not depend on the smoothness of  $H$ , by approximation, we may assume that  $H(y, p)$  is smooth and strictly convex in  $p$  variable (i.e., Tonelli Hamiltonian). In one dimension,  $\bar{H}$  possesses an explicit formula. In particular, we have that  $\bar{H} \in C^1(\mathbb{R})$  (see [6] for instance).

Thanks to (1.2) and Corollary 2.2, Theorem 1.3 follows immediately from the following lemma.

**Lemma 4.1.** *For  $p \in \mathbb{R}$ , let  $v$  be a viscosity solution to*

$$H(y, p + v') = \bar{H}(p) \quad \text{in } \mathbb{T}.$$

*Then any backward characteristic  $\xi : (-\infty, 0] \rightarrow \mathbb{R}$  of  $v$  satisfies that*

$$\left| \frac{\xi(t) - \xi(0)}{t} - \bar{H}'(p) \right| \leq \frac{1}{|t|} \quad \text{for all } t < 0.$$

*Proof.* Fix  $p \in \mathbb{R}$ . There are two cases to be considered.

**Case 1.**  $\bar{H}(p) = \min \bar{H}$ . Then  $\bar{H}'(p) = 0$ .

Let  $\xi$  be a backward characteristic of  $v$  with  $\xi(0) = 0$ . Since  $\xi$  cannot intersect itself, we have either  $\xi((-\infty, 0]) \subset [0, \infty)$  or  $\xi((-\infty, 0]) \subset (-\infty, 0]$ . Without loss of generality, we assume that  $\xi((-\infty, 0]) \subset [0, \infty)$ . Note that,  $\xi$  satisfies

$$\dot{\xi} = H_p(\xi, p + v'(\xi)) \quad \text{for all } t \leq 0.$$

We claim that

$$\xi((-\infty, 0]) \subset [0, 1]. \tag{4.1}$$

Assume otherwise that (4.1) does not hold. Then  $\xi(T) = 1$  for some  $T < 0$ , and we deduce also that  $v \in C^1(\mathbb{T})$ . By periodicity,  $\xi(mT) = m$  for all  $m \in \mathbb{N}$ . Therefore,

$$\lim_{t \rightarrow -\infty} \frac{\xi(t)}{t} = \frac{1}{T} \neq 0 = \bar{H}'(p),$$

which is a contradiction. Thus, (4.1) holds.

**Case 2.**  $\overline{H}(p) > \min \overline{H}$ . Without loss of generality, we assume  $\overline{H}'(p) > 0$ . Let  $\xi$  be a backward characteristic of  $v$  with  $\xi(0) = 0$ . Then  $\xi((-\infty, 0]) \subset (-\infty, 0]$ ,  $v \in C^{1,1}(\mathbb{R})$  and

$$\dot{\xi} = H_p(\xi, p + v'(\xi)) > 0 \quad \text{for all } t \leq 0.$$

Then, by changing of variables  $x = \xi(s)$ , we imply

$$|t| = \int_t^0 ds = \int_t^0 \frac{\dot{\xi}(s)}{\xi(s)} ds = \int_{\xi(t)}^0 \frac{1}{F_1(x)} dx,$$

where  $F_1(x) = H_p(x, p + v'(x))$  for  $x \in \mathbb{R}$ . Accordingly, for  $t < 0$ ,

$$\frac{t}{\xi(t)} = \frac{1}{|\xi(t)|} \int_{\xi(t)}^0 \frac{1}{F_1(x)} dx = \left(1 + \frac{E_t}{\xi(t)}\right) \int_0^1 \frac{1}{F_1(x)} dx,$$

where  $E_t$  is an error term satisfying  $|E_t| \leq 1$  thanks to Lemma 4.2 below. Then

$$\left| \frac{\xi(t)}{t} - \left( \int_0^1 \frac{1}{F_1(x)} dx \right)^{-1} \right| = \frac{|E_t|}{|t|} \leq \frac{1}{|t|}.$$

The proof is complete and we get in addition that

$$\overline{H}'(p) = \left( \int_0^1 \frac{1}{F_1(x)} dx \right)^{-1}.$$

□

□

**Lemma 4.2.** *Assume that  $f \in C(\mathbb{T}, [0, \infty))$  and  $L > 0$  are given. Then*

$$\left| \int_0^L f dy - L \int_0^1 f dy \right| \leq \int_0^1 f dy.$$

We can view this lemma as a quantitative version of the ergodic theorem for periodic functions in one dimension.

*Proof.* For a given real number  $s \in \mathbb{R}$ , denote by  $[s]$  its integer part. We have

$$\begin{aligned} & \left| \int_0^L f dy - L \int_0^1 f dy \right| = \left| \int_0^{[L]} f dy + \int_{[L]}^L f dy - L \int_0^1 f dy \right| \\ &= \left| ([L] - L) \int_0^1 f dy + \int_{[L]}^L f dy \right| \\ &\leq \max \left\{ (L - [L]) \int_0^1 f dy, \int_{[L]}^L f dy \right\} \leq \int_0^1 f dy. \end{aligned}$$

□

**4.1. An example on optimal rate  $O(\varepsilon)$  in one dimension.** We now show that  $O(\varepsilon)$  is indeed the optimal rate of convergence via the following simple proposition.

**Proposition 4.3.** *Assume that  $n = 1$  and  $H(y, p) = \frac{p^2}{2} + V(y)$  for some given  $V \in C(\mathbb{T})$  with  $\max_{\mathbb{T}} V = 0$  and  $V \leq -1$  in  $[-3^{-1}, 3^{-1}]$ . Assume further that  $g \equiv 0$ . For  $\varepsilon > 0$ , let  $u^\varepsilon$  be the solution to  $(C)_\varepsilon$ . Let  $u$  be the solution to  $(C)$ . Then,  $u^\varepsilon$  converges locally uniformly to  $u \equiv 0$  on  $\mathbb{R} \times [0, \infty)$  as  $\varepsilon \rightarrow 0$ . Furthermore, for  $\varepsilon \in (0, 1)$ ,*

$$u^\varepsilon(0, 1) \geq \frac{\varepsilon}{6}. \quad (4.2)$$

*Proof.* It is obvious that  $u^\varepsilon$  converges locally uniformly to  $u \equiv 0$  on  $\mathbb{R} \times [0, \infty)$  as  $\varepsilon \rightarrow 0$ . We just need to prove (4.2). Thanks to the optimal control formula,

$$u^\varepsilon(0, 1) = \inf \left\{ \varepsilon \int_0^{\varepsilon^{-1}} \frac{|\dot{\eta}|^2}{2} - V(\eta) dt : \eta \in \text{AC}([0, \varepsilon^{-1}]), \eta(0) = 0 \right\}.$$

Pick  $\eta \in \text{AC}([0, \varepsilon^{-1}])$  with  $\eta(0) = 0$ . There are two cases to be considered.

Firstly, if  $\eta([0, 3^{-1}]) \subset [-3^{-1}, 3^{-1}]$ , then

$$\varepsilon \int_0^{\varepsilon^{-1}} \frac{|\dot{\eta}|^2}{2} - V(\eta) dt \geq \varepsilon \int_0^{3^{-1}} -V(\eta) dt \geq \frac{\varepsilon}{3}.$$

If not, then without loss of generality, we may assume that there exists  $t \in (0, 3^{-1})$  such that  $\eta(t) = 3^{-1}$ . Then

$$\varepsilon \int_0^{\varepsilon^{-1}} \frac{|\dot{\eta}|^2}{2} - V(\eta) dt \geq \varepsilon \int_0^t \frac{|\dot{\eta}|^2}{2} dt \geq \frac{\varepsilon}{2t} \left( \int_0^t \dot{\eta} dt \right)^2 \geq \frac{\varepsilon}{6}.$$

The proof is complete.  $\square$

## 5. LACK OF CONTINUOUS SELECTION OF $v(y, p)$ WITH RESPECT TO $p$ : WHY THE PDE METHOD IN [9] CANNOT BE EXTENDED

In [9], the authors introduce a rigorous way to approximate the two-scale asymptotic expansion (1.1) by using the ergodic problems and their discounted approximations. More precisely, for each  $p \in \mathbb{R}^n$ , instead of using  $v(y, p)$  directly, they use  $v^\lambda(y, p)$ , solution to the following discount problem

$$\lambda v^\lambda + H(y, p + Dv^\lambda) = 0 \quad \text{in } \mathbb{T}^n,$$

and approximate  $Du(x, t)$  by  $\frac{x-y}{\varepsilon^\beta}$  in (1.1) by using the well-known doubling of variables method. By optimizing choices of  $\lambda$  and  $\beta$  in their arguments,  $O(\varepsilon^{1/3})$  is the best convergence rate obtained.

In order to improve the convergence rate by modifying this approach, it is necessary to have a nice selection of viscosity solutions to the cell problem  $(E)_p$  with respect to  $p$ . This allows us to use directly  $v(y, p)$  instead of  $v^\lambda(y, p)$  in (1.1). For example, if we are able to choose for each  $p \in \mathbb{R}^n$  a solution  $v(y, p)$  of  $(E)_p$  such that

$$p \mapsto v(y, p) \quad \text{is Lipschitz,}$$

then the convergence rate can be improved from  $O(\varepsilon^{1/3})$  to  $O(\varepsilon^{1/2})$ .

In this section, we present an example where there is no continuous selection of  $v(y, p)$  with respect to  $p$ . Assume that  $n = 3$ . Define

$$\begin{cases} L_1 = \{(t, 1/2, 0) : 0 \leq t \leq 1\}, \\ L_2 = \{(0, t, 1/2) : 0 \leq t \leq 1\}, \\ L_3 = \{(1/2, 0, t) : 0 \leq t \leq 1\}. \end{cases}$$

We consider the following mechanical Hamiltonian:

$$\begin{cases} H(y, p) = \frac{1}{2}|p|^2 + V(y) & \text{where } V \in C^\infty(\mathbb{T}^3), \\ \max_{\mathbb{T}^3} V = 0 & \text{and } \{V = 0\} = L_1 \cup L_2 \cup L_3. \end{cases} \quad (5.1)$$

Denote by  $e_1 = (1, 0, 0)$ ,  $e_2 = (0, 1, 0)$  and  $e_3 = (0, 0, 1)$ . For each  $p \in \mathbb{R}^3$ , let  $\mathcal{A}_p$  be the projected Aubry set associated with  $(E)_p$ . We have first the following result.

**Theorem 5.1.** *Assume that (5.1) holds. There exists  $\delta > 0$  sufficiently small such that*

$$\mathcal{A}_{\varepsilon e_i} = L_i \quad \text{for all } \varepsilon \in (0, \delta), i \in \{1, 2, 3\}.$$

The above theorem follows from the following stronger lemma and the constructions of critical subsolutions in its proof.

**Lemma 5.2.** *There exists  $\delta > 0$  sufficiently small such that, for  $|p| < \delta$ ,*

$$\overline{H}(p) = \max \left\{ \frac{|p_1|^2}{2}, \frac{|p_2|^2}{2}, \frac{|p_3|^2}{2} \right\},$$

where  $p = (p_1, p_2, p_3)$ .

*Proof.* We have the following inf-max representation formula for  $\overline{H}(p)$

$$\overline{H}(p) = \inf_{\phi \in C^1(\mathbb{T}^3)} \max_{y \in \mathbb{T}^3} \left( \frac{1}{2} |p + D\phi(y)|^2 + V(y) \right). \quad (5.2)$$

For any  $\phi \in C^1(\mathbb{T}^3)$ , we have  $\phi(0, 1/2, 0) = \phi(1, 1/2, 0)$ . By the Rolle theorem, there exists  $s \in (0, 1)$  such that  $\phi_{y_1}(s, 1/2, 0) = 0$ . Plug this into (5.2) and use the fact that  $V(s, 1/2, 0) = 0$  to get that

$$\overline{H}(p) \geq \frac{|p_1|^2}{2}.$$

By repeating this simple observation two more times, we get, for all  $p \in \mathbb{R}^3$ ,

$$\overline{H}(p) \geq \max \left\{ \frac{|p_1|^2}{2}, \frac{|p_2|^2}{2}, \frac{|p_3|^2}{2} \right\}. \quad (5.3)$$

We now need to obtain the reverse inequality for  $|p|$  small. For  $1 \leq i \leq 3$ , set

$$L_{i,1/8} = \left\{ x \in \mathbb{R}^3 : \text{dist}(x, L_i) \leq \frac{1}{8} \right\}.$$

For  $p \in \mathbb{R}^3$  such that  $|p| = 1$ , we construct a smooth periodic function  $\eta$  satisfying

$$\begin{cases} \eta(y) = -p_2y_2 - p_3y_3 & \text{for } y \in L_{1,1/8}, \\ \eta(y) = -p_1y_1 - p_3y_3 & \text{for } y \in L_{2,1/8}, \\ \eta(y) = -p_1y_1 - p_2y_2 & \text{for } y \in L_{3,1/8}. \end{cases}$$

Set  $C_1 = \max_{\mathbb{R}^3} |D\eta|$ . Choose  $\delta$  such that

$$\delta = \frac{1}{\sqrt{1 + C_1^2}} \left( - \max_{\mathbb{T}^3 \setminus \cup_{i=1}^3 L_{i,1/8}} V \right)^{1/2}.$$

Then for  $\varepsilon \in (0, \delta)$ , we have that  $\phi^\varepsilon = \varepsilon\eta$  is a subsolution to

$$\frac{1}{2}|\varepsilon p + D\phi^\varepsilon|^2 + V = \max \left\{ \frac{|\varepsilon p_1|^2}{2}, \frac{|\varepsilon p_2|^2}{2}, \frac{|\varepsilon p_3|^2}{2} \right\} \quad \text{in } \mathbb{T}^3. \quad (5.4)$$

We combine (5.3) and (5.4) to yield that

$$\overline{H}(p) = \max \left\{ \frac{|p_1|^2}{2}, \frac{|p_2|^2}{2}, \frac{|p_3|^2}{2} \right\} \quad \text{for all } |p| < \delta.$$

Furthermore, for  $i \in \{1, 2, 3\}$  fixed and  $p = \varepsilon e_i$ ,  $\phi_\varepsilon$  is a strict subsolution of (5.4) on  $\mathbb{T}^3 \setminus L_i$ , i.e.,

$$\begin{cases} \frac{1}{2}|\varepsilon p + D\phi^\varepsilon|^2 + V < \overline{H}(p) & \text{in } \mathbb{T}^3 \setminus L_i \\ \frac{1}{2}|\varepsilon p + D\phi^\varepsilon|^2 + V = \overline{H}(p) & \text{in } L_i. \end{cases}$$

Also,  $p + D\phi^\varepsilon = p_i e_i$  along  $L_i$ . We hence deduce easily that  $\mathcal{A}_{\varepsilon e_i} = L_i$  for all  $\varepsilon \in (0, \delta)$ .  $\square$

Note that for this example, the Mather set and Aubry set coincide. It is well-known (see [13]) that solutions to the cell problem  $(E)_p$  are determined by their values on  $\mathcal{A}_p$ . Moreover, all subsolutions are differentiable on  $\mathcal{A}_p$  with the same gradients. Accordingly, for  $p = \varepsilon e_i$ , if  $v(x, p)$  is a viscosity solution to  $(E)_p$ , then

$$Dv = D\phi^\varepsilon \quad \text{on } L_i,$$

which yields that  $v$  is constant along  $L_i$ . An immediate corollary is

**Corollary 5.3.** *For  $\varepsilon \in (0, \delta)$  and  $i \in \{1, 2, 3\}$ , the cell problem  $(E)_{\varepsilon e_i}$  has a unique solution (up to additive constants) which is given by*

$$v(y, \varepsilon e_i) = \inf_{\substack{t > 0 \\ \xi \in \Gamma_{i,t,y}}} \left( \int_0^t \frac{1}{2} |\dot{\xi}|^2 - \dot{\xi} \cdot \varepsilon e_i + V(\xi) + \overline{H}(\varepsilon e_i) ds \right) + c, \quad (5.5)$$

for some constant  $c \in \mathbb{R}$  with  $v = c$  along  $L_i$ . Here  $\Gamma_{i,t,y}$  is the collection of all absolutely continuous curves from  $[0, t]$  to  $\mathbb{R}^3$  such that  $\xi(0) \in L_i + \mathbb{Z}^3$  and  $\xi(t) = y$ .

See also [13] for formula (5.5). Here is the main result of this section.

**Theorem 5.4.** *Assume that (5.1) holds. Then there is no way to select for each  $p \in \mathbb{R}^3$ ,  $v(y, p) \in \text{Lip}(\mathbb{T}^3)$ , a solution to  $(E)_p$ , such that*

$$p \mapsto v(y, p) \text{ is continuous at } p = 0. \quad (5.6)$$

*Proof.* Assume by contradiction that there is a way to select  $v(y, p) \in \text{Lip}(\mathbb{T}^3)$  such that (5.6) holds. Then,

$$\lim_{\varepsilon \rightarrow 0} v(y, \varepsilon e_1) = \lim_{\varepsilon \rightarrow 0} v(y, \varepsilon e_2) = \lim_{\varepsilon \rightarrow 0} v(y, \varepsilon e_3). \quad (5.7)$$

Assume that for  $i \in \{1, 2, 3\}$

$$v(y, \varepsilon e_i) = c_{i\varepsilon} \quad \text{along } L_i.$$

for some constant  $c_{i\varepsilon}$ . We claim that for  $i \in \{1, 2, 3\}$ ,

$$\lim_{\varepsilon \rightarrow 0} v(y, \varepsilon e_i) = d(y, L_i) + c_i, \quad (5.8)$$

where  $c_i = \lim_{\varepsilon \rightarrow 0} c_{i\varepsilon}$ , and

$$d(y, L_i) = \inf_{\substack{t > 0 \\ \xi \in \Gamma_{i,t,y}}} \left( \int_0^t \frac{1}{2} |\dot{\xi}|^2 - V(\xi) ds \right).$$

It suffices to prove this for  $i = 1$  as proofs for  $i = 2, 3$  are similar. It is easy to see that

$$\limsup_{\varepsilon \rightarrow 0} v(y, \varepsilon e_1) \leq d(y, L_1) + c_1.$$

The other direction is more subtle since  $t$  in formula (5.5) (and also  $|\xi(0)|$ ) could go to  $+\infty$ . We need to make use of the special structure of  $V$  and the formula of  $\overline{H}$  here. Note that

$$v(y, \varepsilon e_1) = \inf_{\substack{t > 0 \\ \xi \in \Gamma_{i,t,y}}} \left( \int_0^t \frac{1}{2} |\dot{\xi}|^2 - \dot{\xi} \cdot \varepsilon e_1 - V(\xi) + \overline{H}(\varepsilon e_1) ds \right) + c_{1\varepsilon} \geq c_{1\varepsilon}.$$

Therefore, if  $y \in L_1 + \mathbb{Z}^3$ , then

$$d(y, L_1) + c_1 = c_1 \leq \liminf_{\varepsilon \rightarrow 0} v(y, \varepsilon e_1).$$

We hence only need to consider the case  $y \notin L_1 + \mathbb{Z}^3$ . Fix  $0 < \delta < \frac{1}{8} \text{dist}(y, L_1 + \mathbb{Z}^3)$ . For  $\varepsilon > 0$ , choose  $\xi_\varepsilon \in \Gamma_{i,t_\varepsilon,y}$  for some  $t_\varepsilon > 0$  such that

$$v(y, \varepsilon e_1) \geq \int_0^{t_\varepsilon} \left( \frac{1}{2} |\dot{\xi}_\varepsilon|^2 - \dot{\xi}_\varepsilon \cdot \varepsilon e_1 - V(\xi_\varepsilon) + \overline{H}(\varepsilon e_1) \right) ds + c_{1\varepsilon} - \varepsilon.$$

Note that for  $t_1, t_2 \in [0, t_\varepsilon]$  with  $t_1 < t_2$ ,

$$\begin{aligned} & \int_{t_1}^{t_2} \left( \frac{1}{2} |\dot{\xi}_\varepsilon|^2 - \dot{\xi}_\varepsilon \cdot \varepsilon e_1 - V(\xi_\varepsilon) + \overline{H}(\varepsilon e_1) \right) ds \\ & \geq \frac{t_2 - t_1}{2} \left| \frac{\xi_\varepsilon(t_2) - \xi_\varepsilon(t_1)}{t_2 - t_1} - \varepsilon e_1 \right|^2 - \int_{t_1}^{t_2} V(\xi_\varepsilon) ds \end{aligned} \quad (5.9)$$

For  $1 \leq i \leq 3$ , denote by  $L_{i,\delta}$  the cylinder around  $L_i$  with radius  $\delta$ , that is,

$$L_{i,\delta} = \{x \in \mathbb{R}^3 : \text{dist}(x, L_i) \leq \delta\}.$$

Since  $v(y, \varepsilon e_1)$  is uniformly bounded for all  $\varepsilon \in [0, 1]$ , due to (5.9), for our fixed  $\delta$ , we have that

- the total amount of time that  $\xi_\varepsilon$  stays outside of the tubes  $U = \bigcup_{i=1}^3 (L_{i,\delta} + \mathbb{Z}^3)$  is bounded by a constant  $T_\delta > 0$ , which is independent of  $\varepsilon$ ;

- the number of times that  $\xi_\varepsilon$  jumps between different tubes in  $U$ , and the total jumping distance are bounded by a constant  $N_\delta > 0$ , which is independent of  $\varepsilon$ . Here, a jump of  $\xi_\varepsilon$  refers to an interval  $[t_1, t_2] \subset [0, t_\varepsilon]$  such that  $\xi_\varepsilon(t_1)$  and  $\xi_\varepsilon(t_2)$  belong to two different tubes in  $U$ , and  $\xi_\varepsilon((t_1, t_2)) \cap U = \emptyset$ .

This situation is in fact quite similar to the Hedlund example. Now, denote by  $\bar{t}_{\varepsilon, \delta}$  the exit time of  $\xi_\varepsilon$  from  $L_{1, \delta} + \mathbb{Z}^3$ , that is,

$$\xi_\varepsilon(\bar{t}_{\varepsilon, \delta}) \in (L_{1, \delta} + \mathbb{Z}^3) \quad \text{and} \quad \xi_\varepsilon((\bar{t}_{\varepsilon, \delta}, t_\varepsilon]) \cap (L_{1, \delta} + \mathbb{Z}^3) = \emptyset.$$

Thanks to the observations above, we have that

$$|e_1 \cdot (\xi_\varepsilon(\bar{t}_{\varepsilon, \delta}) - y)| \leq C_\delta,$$

for a constant  $C_\delta$  independent of  $\varepsilon$ . Then,

$$\begin{aligned} v(y, \varepsilon e_1) &\geq \int_{\bar{t}_{\varepsilon, \delta}}^{t_\varepsilon} \left( \frac{1}{2} |\dot{\xi}_\varepsilon|^2 - \dot{\xi}_\varepsilon \cdot \varepsilon e_1 - V(\xi_\varepsilon) + \overline{H}(\varepsilon e_1) \right) ds + c_{1\varepsilon} - \varepsilon \\ &\geq d(y, L_1) - C_\delta - C_\delta \varepsilon + c_{1\varepsilon} - \varepsilon. \end{aligned}$$

By first sending  $\varepsilon \rightarrow 0$ , and then  $\delta \rightarrow 0$ , we have that

$$\liminf_{\varepsilon \rightarrow 0} v(y, \varepsilon e_1) \geq d(y, L_1) + c_1.$$

Therefore, our claim (5.8) holds.

It is clear that  $d(y, L_i) = 0$  for  $y \in L_i + \mathbb{Z}^3$ , and  $d(y, L_i) > 0$  for  $y \in \mathbb{T}^3 \setminus (L_i + \mathbb{Z}^3)$ . Combining (5.7) and (5.8) to get

$$d(y, L_1) + c_1 = d(y, L_2) + c_2 = d(y, L_3) + c_3 \quad \text{for all } y \in \mathbb{T}^3.$$

Without loss of generality, we assume that  $c_1 \geq c_2$ . Then, for  $y \in L_2$ ,

$$c_2 = d(y, L_2) + c_2 = d(y, L_1) + c_1 > c_1,$$

which is a contradiction.  $\square$

**Remark 7.** Note that in this example we pointed out here,  $\{V = 0\} = L_1 \cup L_2 \cup L_3$  is not compactly supported in  $\mathbb{T}^3$  and is of a non-trapping situation, which is why we have the representation formula of  $\overline{H}$  in Lemma 5.2.

Because of a technical reason that we need  $L_{1, 1/8}, L_{2, 1/8}, L_{3, 1/8}$  to be disjoint, we need to work in three dimensional space. In fact, any dimension greater than or equal to three will work. We do not know yet how to construct this kind of example in two dimensions.

**Remark 8.** By refining our arguments, we can actually show that there does not exist continuous selection of supersolutions to  $(E)_p$  either. Indeed, assume otherwise that there is a way to select  $w(y, p) \in \text{Lip}(\mathbb{T}^3)$ , a supersolution to  $(E)_p$ , such that  $p \mapsto w(y, p)$  is continuous at  $p = 0$ . Denote by

$$c_{1\varepsilon} = \min_{y \in L_1} w(y, \varepsilon e_1) = w(y_\varepsilon, \varepsilon e_1) \quad \text{for some } y_\varepsilon \in L_1.$$

By passing to a subsequence if necessary, we assume further that  $\lim_{\varepsilon \rightarrow 0} c_{1\varepsilon} = c_1$ , and  $\lim_{\varepsilon \rightarrow 0} y_\varepsilon = y_0 \in L_1$ . By the comparison principle on Aubry set  $L_1$ , we have



$w(y, \varepsilon e_1) \geq v(y, \varepsilon e_1)$ , where  $v(y, \varepsilon e_1)$  is given in formula (5.5) with  $c = c_{1\varepsilon}$ . Let  $\varepsilon \rightarrow 0$  to yield

$$w(y, 0) \geq d(y, L_1) + c_1 \quad \text{and} \quad w(y_0, 0) = c_1.$$

In particular,

$$\min_{y \in \mathbb{T}^3} w(y, 0) = \min_{y \in L_1} w(y, 0) = c_1 \quad \text{and} \quad w(y, 0) > c_1 \quad \text{for } y \in L_2 \cup L_3.$$

Repeat the above argument for  $c_{2\varepsilon} = \min_{y \in L_2} w(y, \varepsilon e_2)$  to yield the contradiction.

Hence, the method in [9] cannot be employed to obtain an improved upper bound of  $u^\varepsilon - u$ .

## 6. APPENDIX

**6.1. Circle homeomorphism.** We first present a result on circle homeomorphism. A continuous function  $f : \mathbb{R} \rightarrow \mathbb{R}$  is called a *circle homeomorphism* if  $f$  is strictly increasing and

$$f(x + 1) = f(x) + 1 \quad \text{for all } x \in \mathbb{R}.$$

Then, it is well-known that the Poincaré rotation number

$$\beta_f = \lim_{i \rightarrow \infty} \frac{f^i(x)}{i}.$$

exists and is independent of  $x \in \mathbb{R}$ . Moreover,

$$|f^i(x) - f(x) - i\beta_f| \leq 1 \quad \text{for all } i \in \mathbb{Z}. \tag{6.1}$$

Also,  $\beta_f = \frac{p}{q} \in \mathbb{Q}$  with  $p \in \mathbb{Z}, q \in \mathbb{N}$  if and only if there exists  $x_0 \in \mathbb{R}$  such that

$$f^q(x_0) = f(x_0) + p.$$

Here, for  $i \in \mathbb{N}$ ,  $f^i$  represents the  $i$ -th iteration of  $f$ . See [2, Chapter 3, §11] for details.

**6.2. Rotation vectors in two dimensions.** Assume  $n = 2$ . Let  $H$  be a smooth Hamiltonian satisfying (H1)–(H3) and furthermore that  $p \mapsto H(y, p)$  is strictly convex for each  $y \in \mathbb{T}^2$ . Let  $L$  be the corresponding Lagrangian.

A Lipschitz continuous curve  $\gamma : \mathbb{R} \rightarrow \mathbb{R}^2$  is called an *absolute minimizer* associated with  $L + c$  for some  $c \in \mathbb{R}$  if

$$\int_{t_1}^{t_2} L(\gamma, \dot{\gamma}) + c ds \leq \int_{s_1}^{s_2} L(\delta, \dot{\delta}) + c ds$$

for any  $t_1 < t_2, s_1 < s_2$  and Lipschitz continuous curve  $\delta$  satisfying  $\delta(s_i) = \gamma(t_i)$  for  $i = 1, 2$ . Two absolute minimizers associated with  $L + c$  cannot intersect twice unless they are the same after suitable translation in time. This non-crossing property, together with the two dimensional topology, plays a crucial role in the Aubry-Mather theory, which provides detailed information about distributions of absolute minimizers (see [3]). For reader's convenience, let us give a brief explanation about this important non-crossing fact here. Assume otherwise that two distinct (up to

translation) absolute minimizers  $\gamma_1$  and  $\gamma_2$  intersect at least twice. By translation, we may assume that there are  $a, b_1, b_2 \in \mathbb{R}$  such that  $a < b_1 \leq b_2$  and

$$\gamma_1(a) = \gamma_2(a), \quad \gamma_1(b_1) = \gamma_2(b_2).$$

It is clear that

$$\int_a^{b_1} L(\gamma_1, \dot{\gamma}_1) + c \, ds = \int_a^{b_2} L(\gamma_2, \dot{\gamma}_2) + c \, ds.$$

Let  $\gamma_3 : [a, b_1 + 1] \rightarrow \mathbb{R}^2$  be such that

$$\gamma_3(s) = \begin{cases} \gamma_1(s) & \text{for } s \in [a, b_1], \\ \gamma_2(s + b_2 - b_1) & \text{for } s \in [b_1, b_1 + 1]. \end{cases}$$

Since

$$\int_a^{b_2+1} L(\gamma_2, \dot{\gamma}_2) + c \, ds = \int_a^{b_1+1} L(\gamma_3, \dot{\gamma}_3) + c \, ds,$$

$\gamma_3|_{[a, b_1+1]}$  is also a minimizer of the action

$$\int L(\gamma, \dot{\gamma}) + c \, ds$$

with corresponding fixed endpoints  $\gamma_2(a) = \gamma_3(a)$  and  $\gamma_2(b_2 + 1) = \gamma_3(b_1 + 1)$ . Hence  $\gamma_3$  is  $C^2$  and solves the following Euler-Lagrange equations

$$\frac{d}{ds} (D_q L(\gamma_3(s), \dot{\gamma}_3(s))) = D_x L(\gamma_3(s), \dot{\gamma}_3(s)) \quad \text{for all } s \in [a, b_1 + 1].$$

Accordingly, at the junction  $\gamma_1(b_1) = \gamma_2(b_2) = \gamma_3(b_1)$ , we must have that

$$\gamma_1'(b_1) = \gamma_2'(b_2) = \gamma_3'(b_1).$$

Since  $\gamma_1$  and  $\gamma_2$  are also solutions to the above Euler-Lagrange equation, the uniqueness result of second order ODEs yields that  $\gamma_1(t) = \gamma_2(t + b_2 - b_1)$ , which is absurd.

The following theorem was established in [8].

**Theorem 6.1.** *If  $\overline{H}(p) > \min_{\mathbb{R}^2} \overline{H}$ , the set of sub-differentials  $\partial \overline{H}(p)$  is a radical segment, that is,*

$$\partial \overline{H}(p) = \{r n_p : r \in [r_1(p), r_2(p)]\} \quad \text{for some unit vector } n_p \text{ and } r_1(p), r_2(p) > 0.$$

*In particular, this implies that for  $s > \min_{\mathbb{R}^2} \overline{H}$ , the level curve  $\{\overline{H} = s\}$  is  $C^1$ .*

Note that the above is in general false when  $n \geq 3$  (see Lemma 5.2). For readers' convenience, we provide a sketch of the proof. The basic idea is quite simple. Suppose that  $q_1, q_2 \in \partial \overline{H}(p)$ . Then, there exist two Mather measures  $\mu_1$  and  $\mu_2$  associated with  $p$  such that, for  $i = 1, 2$ , the rotation vector

$$\iint_{\mathbb{T}^2 \times \mathbb{R}^2} q \, d\mu_i = q_i.$$

Accordingly, if  $q_1$  and  $q_2$  are not parallel, then there are two different orbits from supports of  $\mu_1$  and  $\mu_2$  respectively, which intersect each other. This is impossible.

This conclusion also implies the existence of  $s_p \in [-\infty, \infty]$  such that, for any orbit  $\xi(t) = (x(t), y(t))$  on the Mather set  $\widetilde{\mathcal{M}}_p$ ,

$$\lim_{|t| \rightarrow \infty} \frac{y(t)}{x(t)} = s_p.$$

**6.3. Identification with circle homeomorphisms.** Fix  $p \in \mathbb{R}^2$ . Let  $\widetilde{\mathcal{M}}_p$  be the Mather set corresponding to  $p$ .

Assume that  $\xi : \mathbb{R} \mapsto \mathbb{T}^2$  is an orbit on  $\widetilde{\mathcal{M}}_p$ , i.e.,

$$\{(\xi(t), \dot{\xi}(t)) : t \in \mathbb{R}\} \subset \widetilde{\mathcal{M}}_p.$$

We lift  $\xi$  to  $\mathbb{R}^2$  and still write it the same by abuse of notation. Let us note right away that orbits on  $\widetilde{\mathcal{M}}_p$  do not intersect with each other since they are all absolute minimizers of  $\int L(\gamma, \dot{\gamma}) + \overline{H}(p) ds$ . Therefore, they are totally ordered in  $\mathbb{R}^2$  (see Figure 6.1).

In the following, we explain how to associate  $\xi = (x(t), y(t))$  with a circle map  $f : \mathbb{R} \rightarrow \mathbb{R}$  when  $\overline{H}(p) > \min_{\mathbb{R}^2} \overline{H}$ , which is well-known in the Aubry-Mather theory. See [3, Chapter 3 and 6] for details from geometric point of view. Without loss of generality, we assume that

$$\lim_{|t| \rightarrow \infty} \frac{y(t)}{x(t)} = s_p \in [0, 1].$$

Let  $\eta : \mathbb{R} \mapsto \mathbb{T}^2$  be a periodic trajectory associated with another Mather set on the same energy level with rotation vector parallel to  $(0, 1)$ . More precisely, there exists  $p' \in \mathbb{R}^2$  and  $T > 0$  such that  $\overline{H}(p') = \overline{H}(p)$ ,

$$\{(\eta(t), \dot{\eta}(t)) : t \in \mathbb{R}\} \subset \widetilde{\mathcal{M}}_{p'},$$

and

$$\eta(T) - \eta(0) = (0, 1).$$

The existence of periodic orbits  $\eta$  with rational rotation vector is well-known (see [3]). Clearly, for each  $k \in \mathbb{Z}$ ,  $\xi$  intersects with  $\eta_k = \eta + (k, 0)$  exactly once since both  $\xi$  and  $\eta$  are absolute minimizers of the action

$$\int L(\gamma, \dot{\gamma}) + \overline{H}(p) ds.$$

Without loss of generality, we may assume that

$$\xi(0) = \eta(0) \in [0, 1]^2.$$

For each  $k \in \mathbb{Z}$ , let  $a_k \in \mathbb{R}$  be such that

$$\xi \cap \eta_k = \eta_k(a_k T).$$

Since orbits on  $\widetilde{\mathcal{M}}_p$  are totally ordered in  $\mathbb{R}^2$ , either  $a_k = 0$  for all  $k \in \mathbb{Z}$  or  $\{a_k\}_{k \in \mathbb{Z}}$  is a strictly increasing sequence. Moreover, for fixed  $k > l$ ,

$$a_k - a_l = i \in \mathbb{Z} \implies a_{k+m} - a_{l+m} = i \quad \text{for all } m \in \mathbb{Z}.$$

Indeed,  $a_k - a_l = i$  means that there exists  $\alpha \in \mathbb{R}$  such that  $\xi(s + \alpha) = \xi(s) + (k - l, i)$  for all  $s \in \mathbb{R}$ , and hence, the implication follows.

Thus, there exists a circle homeomorphism  $f$  such that

$$f(a_k) = a_{k+1} \quad \text{for all } k \in \mathbb{Z}.$$

See [3, Theorem 3.15] for further details on the definition of  $f$ .

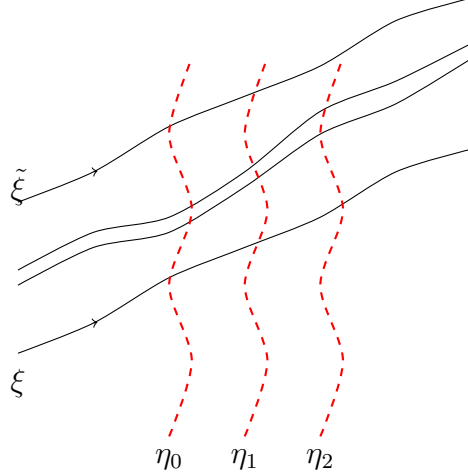


FIGURE 6.1. Orbits in  $\widetilde{\mathcal{M}}_p$  and  $\{\eta_k\}_{k \in \mathbb{Z}}$

**6.4. Uniform convergence of slope of orbits on Mather sets.** Through suitable translations, we may assume that

$$\min_{\mathbb{R}^2} \overline{H} = \overline{H}(0) = 0.$$

Let  $p, p', \xi$  and  $\eta$  be from the previous section. Denote  $0 < r = \overline{H}(p) = \overline{H}(p')$ ,

$$S_r = \{\overline{H} = r\}, \quad a_r = \max_{\tilde{p} \in S_r} |\tilde{p}|, \quad b_r = \max \{|D_p H(y, \tilde{p})| : H(y, \tilde{p}) = r\},$$

and

$$d = \max \{|\tilde{p}| : H(y, \tilde{p}) = 0\}.$$

Let  $v$  and  $v'$  be viscosity solutions to

$$H(y, p + Dv) = \overline{H}(p) = r \quad \text{in } \mathbb{T}^2$$

and

$$H(y, p' + Dv') = \overline{H}(p') = r \quad \text{in } \mathbb{T}^2,$$

respectively. Note that  $v, v'$  are smooth along  $\xi, \eta$ , respectively, and

$$\dot{\xi} = D_p H(\xi, p + Dv(\xi)) \quad \text{and} \quad \dot{\eta} = D_p H(\eta, p' + Dv'(\eta)).$$

Therefore,

$$|\dot{\xi}|, |\dot{\eta}| \leq b_r. \tag{6.2}$$

Also, for any Lipschitz continuous curve  $\gamma : [s_1, s_2] \rightarrow \mathbb{R}^2$ ,

$$\begin{aligned} \int_{s_1}^{s_2} L(\gamma, \dot{\gamma}) ds &= \int_{s_1}^{s_2} L(\gamma, \dot{\gamma}) + H(\gamma, Dw) ds \\ &\geq \int_{s_1}^{s_2} \dot{\gamma} \cdot Dw ds = w(\gamma(s_2)) - w(\gamma(s_1)) \geq -d\sqrt{2}. \end{aligned} \quad (6.3)$$

Here  $w$  is a viscosity solution to (E)<sub>0</sub>. Although  $w$  might not be differentiable along  $\gamma$ , the above calculation can be easily justified by using standard convolutions of  $w$ .

Next we give uniform lower and upper bounds of the period  $T$  of  $\eta$ . Since the rotation vector

$$\frac{\eta(T) - \eta(0)}{T} = \frac{1}{T}(0, 1) \in \partial \bar{H}(p')$$

and  $p' \cdot q \geq \bar{H}(p') = r$  for any  $q \in \partial \bar{H}(p')$ , we have that

$$\frac{1}{b_r} \leq T \leq \frac{a_r}{r}. \quad (6.4)$$

**Lemma 6.2.** *Assume that  $s_p \in [0, 1]$ . Then for all  $t \in \mathbb{R}$ ,*

$$|y(t) - s_p x(t)| \leq C.$$

Here  $C$  is a constant depending only on  $r, a_r, b_r$  and  $d$ .

*Proof.* Assume that

$$\xi(t_k) = \eta_k(a_k T).$$

In light of (6.2) and (6.4)

$$\left| \eta(t) - \eta(0) - \left(0, \frac{t}{T}\right) \right| \leq C.$$

Therefore,

$$|x(t_k) - k| \leq C \quad \text{and} \quad |y(t_k) - a_k| \leq C.$$

Hence,  $\lim_{k \rightarrow \infty} \frac{a_k}{k} = s_p$ . Since  $t_0 = a_0 = 0$ , thanks to (6.1),

$$|a_k - ks_p| \leq 1.$$

Accordingly,

$$|y(t_k) - s_p x(t_k)| \leq C \quad \text{for all } k \in \mathbb{Z}. \quad (6.5)$$

Next we claim that

$$|t_k - t_{k+1}| \leq C \quad \text{for all } k \in \mathbb{Z}. \quad (6.6)$$

In fact, from the previous calculations, we have that

$$|x(t_{k+1}) - x(t_k)| \leq C \quad \text{and} \quad |y(t_{k+1}) - y(t_k)| \leq C,$$

which, together with (6.3), yields

$$\begin{aligned} C &\geq p \cdot (\xi(t_{k+1}) - \xi(t_k)) + v(\xi(t_{k+1})) - v(\xi(t_k)) = \int_{t_k}^{t_{k+1}} L(\xi, \dot{\xi}) + \bar{H}(p) ds \\ &= r(t_{k+1} - t_k) + \int_{t_k}^{t_{k+1}} L(\xi, \dot{\xi}) ds \\ &\geq r(t_{k+1} - t_k) - d\sqrt{2}. \end{aligned}$$

Thus, (6.6) holds. Combine (6.2), (6.5) and (6.6) to imply

$$|y(t) - s_p x(t)| \leq C \quad \text{for all } t \in \mathbb{R}. \quad (6.7)$$

**Remark 9.** We would like to highlight that constant  $C$  in (6.7) above does not depend on the smoothness of  $H$ . Write

$$\begin{aligned} M_r(x) &= \max \{|p| : H(x, p) = r\}, \\ D_r(x) &= \min \{|p_2 - p_1| : H(x, p_2) = 2r, H(x, p_1) = r\}. \end{aligned}$$

Then it is easy to see that

$$a_r \leq \max_{x \in \mathbb{R}^2} M_r(x) \quad \text{and} \quad b_r \leq \frac{r}{\min_{x \in \mathbb{R}^2} D_r(x)}.$$

□

## REFERENCES

- [1] S. N. Armstrong, P. Cardaliaguet, P. E. Souganidis, *Error estimates and convergence rates for the stochastic homogenization of Hamilton-Jacobi equations*, J. Amer. Math. Soc. 27 (2014), no. 2, 479–540.
- [2] V. I. Arnold, *Geometrical methods in the theory of ordinary differential equations*, Second edition, Grundlehren der mathematischen Wissenschaften, 250, 1988, Springer-Verlag.
- [3] V. Bangert, *Mather Sets for Twist Maps and Geodesics on Tori*, Dynamics Reported, Volume 1, 1988.
- [4] V. Bangert, *Minimal geodesics*, Ergod. Th. & Dynam. Sys. (1989), 10, 263–286.
- [5] V. Bangert, *Geodesic rays, Busemann functions and monotone twist maps*, Calculus of Variations and Partial Differential Equations, January 1994, Volume 2, Issue 1, 49–63.
- [6] P. Bernard, *The asymptotic behaviour of solutions of the forced Burgers equation on the circle*, Nonlinearity 18 (2005) 101–124.
- [7] F. Camilli, A. Cesaroni, C. Marchi, *Homogenization and vanishing viscosity in fully nonlinear elliptic equations: rate of convergence estimates*, Adv. Nonlinear Stud. 11 (2011), no. 2, 405–428.
- [8] M. J. Carneiro, *On minimizing measures of the action of autonomous Lagrangians*, Nonlinearity 8 (1995) 1077–1085.
- [9] I. Capuzzo-Dolcetta, H. Ishii, *On the rate of convergence in homogenization of Hamilton-Jacobi equations*, Indiana Univ. Math. J. 50 (2001), no. 3, 1113–1129.
- [10] W. E, *Aubry-Mather theory and periodic solutions of the forced Burgers equation*, *Comm. Pure Appl. Math.* 52 (1999), no. 7, 811–828.
- [11] L. C. Evans, *Periodic homogenisation of certain fully nonlinear partial differential equations*, Proc. Roy. Soc. Edinburgh Sect. A 120 (1992), no. 3-4, 245–265.
- [12] L. C. Evans, D. Gomes, *Effective Hamiltonians and Averaging for Hamiltonian Dynamics. I*, Arch. Ration. Mech. Anal. 157 (2001), no. 1, 1–33.
- [13] A. Fathi, *Weak KAM Theorem in Lagrangian Dynamics*.
- [14] D. A. Gomes, *Viscosity solutions of Hamilton-Jacobi equations, and asymptotics for Hamiltonian systems*, Calc. Var. 14, 345–357 (2002).
- [15] G. A. Hedlund, *Geodesics on a two-dimensional Riemannian manifold with periodic coefficients*, Ann. of Math. 33 (1932), 719–739.
- [16] P.-L. Lions, G. Papanicolaou and S. R. S. Varadhan, *Homogenization of Hamilton-Jacobi equations*, unpublished work (1987).
- [17] S. Luo, Y. Yu, H. Zhao, *A new approximation for effective Hamiltonians for homogenization of a class of Hamilton-Jacobi equations*, Multiscale Model. Simul. 9 (2011), no. 2, 711–734.

- [18] H. Mitake, H. V. Tran, *Homogenization of weakly coupled systems of Hamilton-Jacobi equations with fast switching rates*, Arch. Ration. Mech. Anal. 211 (2014), no. 3, 733–769.

(H. Mitake) GRADUATE SCHOOL OF MATHEMATICAL SCIENCES, UNIVERSITY OF TOKYO 3-8-1 KOMABA, MEGURO-KU, TOKYO, 153-8914, JAPAN  
*E-mail address:* mitake@ms.u-tokyo.ac.jp

(H. V. Tran) DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WISCONSIN MADISON, VAN VLECK HALL, 480 LINCOLN DRIVE, MADISON, WISCONSIN 53706, USA  
*E-mail address:* hung@math.wisc.edu

(Y. Yu) DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA AT IRVINE, CALIFORNIA 92697, USA  
*E-mail address:* yyu1@math.uci.edu