A course on weak KAM theory

Hung Vinh Tran, Yifeng Yu

(H. V. Tran) Department of Mathematics, University of Wisconsin Madison, Van Vleck hall, 480 Lincoln Drive, Madison, WI 53706, USA

Email address: hung@math.wisc.edu

(Y. Yu) Department of Mathematics, University of California at Irvine, California 92697, USA

Email address: yyu1@math.uci.edu

2020 *Mathematics Subject Classification.* Primary 35-XX;
Secondary 49-XX, 37-XX.

Key words and phrases. weak KAM theory; Hamilton–Jacobi equations; viscosity solutions; Optimal Control Theory; homogenization theory; effective Hamiltonians.

Contents

Preface

In this set of lecture notes, we present the weak Kolmogorov–Arnold–Moser (KAM) theory and its connections to other research areas. The emphasis here is more on dynamics and dynamical methods, and less on partial differential equations. This is the first draft that contains basic materials on the subject. In this current form, the lecture notes should be used for educational purposes only.

We list here the main contents covered in the book.

(1) The Legendre transform and its properties.

(2) Action functionals, existence and regularity of their minimizers.

(3) The weak KAM theorem via both dynamical system and PDE viewpoints.

(4) Invariant measures and sets including Mather measures, Mather sets, and Aubry sets.

(5) Aubry-Mather theory in two dimensions in the smooth setting.

(6) Aubry-Mather theory in the merely continuous setting.

(7) Optimal rate of convergence for periodic homogenization of Hamilton-Jacobi equations in the convex setting.

(8) Large time behavior for Hamilton-Jacobi equations in the torus.

In the appendices, we give some basic points of circle homeomorphisms, and the method of characteristics to solve Hamilton-Jacobi equations locally.

Some parts of the book are based on various topic courses that we have taught at UW-Madison and UC Irvine. We would like to thank Son Tu, who provided us the first draft of the lecture notes of a graduate dynamical system course (Math 807) that Hung Tran taught in Spring 2021 at UW-Madison. We thank Jianxing Du for nice remarks to clarify Theorem 1.14.

Our goal in the very long run is to turn this into a research monograph on weak Kolmogorov–Arnold–Moser (KAM) theory and its connections to other research areas. It is an ongoing process that will take us a long time to finish.

Hung Tran was supported in part by NSF CAREER grant DMS-1843320, a Simons Fellowship, and a Vilas Faculty Early-Career Investigator Award (2022-2025) during the writing of this book. Yifeng Yu was partially supported by NSF grant 2000191 during the writing of this first draft.

Hung Vinh Tran, Yifeng Yu
Fall 2022

Chapter 1

The Legendre transform

1.1. Legendre's transform

Let $H : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Our main goals in this chapter are to study properties of H and its Legendre transform deeply. Generally speaking, *convexity* is *one-sided linearity*. More precisely, H is the supremum of all affine functions whose graphs stay below its graph; that is, there exists an index set $\mathcal{A}$ such that, for $p \in \mathbb{R}^n$,

$$H(p) = \sup \left\{ v_\alpha \cdot p + c_\alpha \,:\, \alpha \in \mathcal{A} \right\},$$

where $\{v_\alpha\}_{\alpha \in \mathcal{A}} \subset \mathbb{R}^n, \{c_\alpha\}_{\alpha \in \mathcal{A}} \subset \mathbb{R}$.

Definition 1.1 (Legendre's transform)**.** Assume $H : \mathbb{R}^n \to \mathbb{R}$ is convex and superlinear, that is,

$$\lim_{|p| \to \infty} \frac{H(p)}{|p|} = +\infty.$$

Then, the Legendre transform of H, $H^* : \mathbb{R}^n \to \mathbb{R}$, is defined as

$$H^*(v) = \sup_{p \in \mathbb{R}^n} \left(p \cdot v - H(p) \right) \qquad \text{for } v \in \mathbb{R}^n.$$

Example 1.2. If $H(p) = \frac{1}{2}|p|^2$ for $p \in \mathbb{R}^n$, then $H^*(v) = \frac{1}{2}|v|^2$ for $v \in \mathbb{R}^n$.

Remark 1.3. Typically, we say that H is the Hamiltonian, and $L = H^*$ is the corresponding Lagrangian.

Let us now explain the geometric meaning of the Legendre transform. Consider all hyperplanes touching the graph of H from below of the form $l_v(p) = p \cdot v + c$. For each fixed vector $v \in \mathbb{R}^n$, the corresponding hyperplane

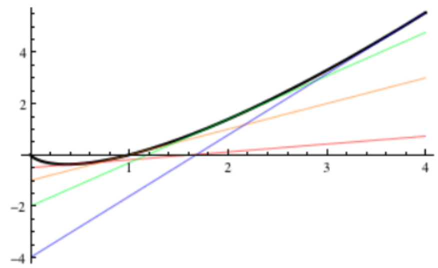

Figure 1. Geometric meaning of the Legendre transform.

$l_v(p) = p \cdot v + c$ touches H from below, which means that at the touching point $p_v \in \mathbb{R}^n$, $H(p_v) = p_v \cdot v + c$, and hence,

$$L(v) = H^*(v) = \sup_{p \in \mathbb{R}^n} \Big(p \cdot v - H(p) \Big) = -c = -l_v(0).$$

Definition 1.4 (Supporting hyperplanes)**.** Assume $H : \mathbb{R}^n \to \mathbb{R}$ is convex and superlinear. For each fixed vector $v \in \mathbb{R}^n$, if the hyperplane $l_v(p) = p \cdot v + c$ touches H from below for a given $c \in \mathbb{R}$, then we say that l_v is a supporting hyperplane of H.

As noted above, if l_v is a supporting hyperplane of H, then

$$L(v) = H^*(v) = -l_v(0). \tag{1.1}$$

1.2. Basic properties of the Legendre transform

1.2.1. Basic properties. We proceed with some first basic properties of the Legendre transform.

Lemma 1.5. *Let $L = H^*$ be the Legendre transform of H. Then, L is finite, convex, and superlinear.*

It is worth noting that if H is not superlinear, then L is still defined, but it could be infinite at some places. For example, if $H(p) = |p|$ for $p \in \mathbb{R}^n$, then

$$L(v) = H^*(v) = \begin{cases} 0 & \text{for } |v| \le 1, \\ +\infty & \text{otherwise.} \end{cases}$$

Proof of Lemma 1.5. Fix $v \in \mathbb{R}^n$. Since H is superlinear in p, we have

$$p \cdot v - H(p) = |p| \left(\frac{p \cdot v}{|p|} - \frac{H(p)}{|p|} \right) \to -\infty \qquad \text{as } |p| \to \infty,$$

which means that

$$L(v) = \sup_{p \in \mathbb{R}^n} \Big(p \cdot v - H(p)\Big) = \max_{p \in \mathbb{R}^n} \Big(p \cdot v - H(p)\Big) < \infty.$$

Thus, L is finite. Besides, $v \mapsto L(v)$ is convex as it is a supremum of a family of affine functions in v.

Now, we prove that L is superlinear in v. For $v \neq 0$, choose $p = s\frac{v}{|v|}$, then for any $s > 0$, we have

$$\begin{aligned} L(v) &= \sup_{p \in \mathbb{R}^n} \Big(p \cdot v - H(p)\Big) \\ &\geq \left(s\frac{v}{|v|}\right) \cdot v - H\left(s\frac{v}{|v|}\right) \geq s|v| - \max_{|p| \leq s} H(p). \end{aligned}$$

Hence, for any fixed $s > 0$,

$$\liminf_{|v| \to \infty} \frac{L(v)}{|v|} \geq s - \limsup_{|v| \to \infty} \left(\frac{1}{|v|} \max_{|p| \leq s} H(p)\right) = s,$$

which yields that

$$\lim_{|v| \to \infty} \frac{L(v)}{|v|} = +\infty.$$

□

Let us now define subgradients of convex functions.

Definition 1.6 (Subgradients of convex functions)**.** Assume $H : \mathbb{R}^n \to \mathbb{R}$ is convex. Fix $p_0 \in \mathbb{R}^n$. The subgradient of H at p_0 is defined as

$$\begin{aligned} \partial H(p_0) &= \{v \in \mathbb{R}^n \,:\, H(p) \geq H(p_0) + v \cdot (p - p_0) \text{ for all } p \in \mathbb{R}^n\} \\ &= \{v \in \mathbb{R}^n \,:\, l_v(p) = p \cdot v + H(p_0) - p_0 \cdot v \\ &\qquad\qquad\qquad\qquad \text{is a supporting hyperplane of } H\}. \end{aligned}$$

We note that $\partial H(p_0) \neq \emptyset$, and if H is differentiable at p_0, then

$$\partial H(p_0) = \{DH(p_0)\}.$$

Lemma 1.7. *If H is convex, then $L^* = (H^*)^* = H$.*

Proof. It is clear that

$$L(v) = \sup_{p \in \mathbb{R}^n} \Big(p \cdot v - H(p)\Big) \geq p \cdot v - H(p) \qquad \text{for any } p \in \mathbb{R}^n.$$

This implies

$$H(p) + L(v) \geq p \cdot v \qquad \text{for all } p, q \in \mathbb{R}^n. \tag{1.2}$$

In particular,

$$H(p) \geq \sup_{v \in \mathbb{R}^n} (p \cdot v - L(v)) = L^*(p).$$

Therefore, $H \geq L^*$. Conversely, we have

$$L^*(p) = \sup_{v\in\mathbb{R}^n} \Big(p \cdot v - L(v)\Big) = \sup_{v\in\mathbb{R}^n} \Big(p \cdot v - \sup_{r\in\mathbb{R}^n} \Big(r \cdot v - H(r)\Big)\Big)$$
$$= \sup_{v\in\mathbb{R}^n} \inf_{r\in\mathbb{R}^n} \Big((p - r) \cdot v + H(r)\Big).$$

Thus

$$L^*(p) \geq \inf_{r\in\mathbb{R}} \Big(H(r) - (r - p) \cdot v\Big) \qquad \text{for all } v \in \mathbb{R}^n.$$

Pick $v \in \partial H(p)$. By the definition of subgradients,

$$H(r) - (r - p) \cdot v \geq H(p) \qquad \text{for all } r \in \mathbb{R}^n.$$

Therefore, $L^* \geq H$. The proof is complete. □

Remark 1.8. An important inequality arises from the proof of Lemma 1.7 is

$$H(p) + L(v) \geq p \cdot v \qquad \text{for all } p, v \in \mathbb{R}^n.$$

This is often called the convex duality inequality or Fenchel's inequality. It is natural to ask when we have equality in the above. From the geometric meaning of the Legendre transform and the definition of subgradients,

$$H(p) + L(v) = p \cdot v \iff p \in \partial L(v) \iff v \in \partial H(p). \tag{1.3}$$

Remark 1.9. We say that the Legendre transform is involutive in the set of convex functions, that is, for H convex,

$$(H^*)^* = H.$$

Moreover, it is clear from the definition of the Legendre transform that, for H, G convex and $H \geq G$,

$$H^* \leq G^*.$$

We say that the Legendre transform reverses the ordering in the set of convex functions.

We now study the differentiability of convex functions and their Legendre transforms.

Theorem 1.10. *Assume that H is convex and differentiable. Then, $H \in C^1(\mathbb{R}^n)$.*

Proof. Assume $\{p_k\} \to p_0 \in \mathbb{R}^n$. We now show that $DH(p_k) \to \xi_0 = DH(p_0)$. There exists $C > 0$ such that $|p_k| \leq C$ for all $k \in \mathbb{N}$. As H is convex,

$$H(p_k + h) \geq H(p_k) + DH(p_k) \cdot h \qquad \text{for all } |h| \leq 1.$$

Thus, $|DH(p_k)| \leq 2\max_{|p|\leq C+1}|H(p)|$ for all $k \in \mathbb{N}$. By passing to a subsequence if necessary, we may assume that $DH(p_k) \to \xi_0$ for some $\xi_0 \in \mathbb{R}^n$. For all $p \in \mathbb{R}^n$,

$$H(p) \geq H(p_k) + DH(p_k) \cdot (p - p_k).$$

Let $k \to \infty$ to deduce that, for $p \in \mathbb{R}^n$,

$$H(p) \geq H(p_0) + \xi_0 \cdot (p - p_0),$$

which gives that $\xi_0 \in \partial H(p_0)$. As H is differentiable, $\xi_0 = DH(p_0)$, and hence, $DH(p_k) \to DH(p_0)$. $\square$

The above proof also implies the following lemma.

Lemma 1.11. *Assume that H is convex. Then, the following properties hold.*

(i) *(Boundedness of subgradients) For each $R > 0$, there exists $C_R > 0$ such that*
$$\partial H(B(0,R)) \subset B(0, C_R).$$
(ii) *(Stability) If $p_k \to p$ and $v_k \in \partial H(p_k)$ such that $v_k \to v$, then $v \in \partial H(p)$.*

1.2.2. Exercises.

Exercise 1. Assume that $H : \mathbb{R}^n \to \mathbb{R}$ is convex. Show that $\partial H(p) \neq \emptyset$ for each $p \in \mathbb{R}^n$.

Exercise 2. If H is not convex, then we can still define H^*. In this situation, how does $H^{**} = (H^*)^*$ relate to H? Give one explicit example of H, and compute H^*, H^{**}.

1.2.3. Strictly convex Hamiltonians.

Theorem 1.12. *Assume that H is convex and superlinear. Then, the following are equivalent.*

(i) *H is strictly convex, that is, for $p_1 \neq p_2$ and $s \in (0,1)$,*

$$H(sp_1 + (1-s)p_2) < sH(p_1) + (1-s)H(p_2). \tag{1.4}$$

(ii) *$\partial H(p_1) \cap \partial H(p_2) = \emptyset$ if $p_1 \neq p_2$.*
(iii) *$L = H^* \in C^1(\mathbb{R}^n)$.*

Proof. We first show that (i) implies (ii). If $\xi \in \partial H(p_1) \cap \partial H(p_2)$ for some $p_1 \neq p_2$, then by definition of subgradients, we have

$$\begin{cases} H(sp_1 + (1-s)p_2) \geq H(p_1) + \xi \cdot (p_2 - p_1)(1-s), \\ H(sp_1 + (1-s)p_2) \geq H(p_2) + \xi \cdot (p_1 - p_2)s, \end{cases}$$

for $s \in (0,1)$. Multiplying the first equation by s, the second equation by $(1-s)$ and adding them together, we obtain

$$H(sp_1 + (1-s)p_2) \geq sH(p_1) + (1-s)H(p_2),$$

which contradicts (i).

Next, we prove that (ii) implies (iii). By Theorem 1.10, it suffices to show that $\partial L(v)$ is a singleton at any $v \in \mathbb{R}^n$. Assume by contradiction that for some $v \in \mathbb{R}^n$, $p_1, p_2 \in \partial L(v)$ for $p_1 \neq p_2$. Then, in light of (1.3),

$$v \in \partial H(p_1) \cap \partial H(p_2) = \emptyset,$$

which is absurd.

Finally, we show (iii) implies (i). Assume otherwise that H is not strictly convex, that is,

$$H(s_0 p_1 + (1-s_0)p_2) = s_0 H(p_1) + (1-s_0)H(p_2).$$

for some $s_0 \in (0,1)$ and $p_1 \neq p_2$. Then, for all $s \in (0,1)$, there holds

$$H(sp_1 + (1-s)p_2) = sH(p_1) + (1-s)H(p_2).$$

Take $v \in \partial H(p_s)$ where $p_s = sp_1 + (1-s)p_2$ for $s \in (0,1)$. Then, for $p \in \mathbb{R}^n$, we have $H(p) - H(p_s) \geq v \cdot (p - p_s)$, which gives that

$$\begin{aligned} H(p) - H(p_1) &\geq H(p_s) - H(p_1) + v \cdot (p - p_s) \\ &= (1-s)\big(H(p_2) - H(p_1)\big) + v \cdot (p - p_s) \\ &\geq (1-s)v \cdot (p_2 - p_1) + v \cdot (p - p_s) \\ &= v \cdot (p - p_1). \end{aligned}$$

Therefore, $v \in \partial H(p_1)$, and similarly $v \in \partial H(p_2)$ as well. This implies $p_1, p_2 \in \partial L(v) = \{DL(v)\}$, which is a contradiction. $\square$

Theorem 1.13. *Assume $H \in C^k(\mathbb{R}^n)$ with $k \geq 2$. Assume further that H is convex, superlinear, and is locally uniformly convex, i.e., $D^2H(p) > 0$ for all $p \in \mathbb{R}^n$. Then,*

- *$L \in C^k(\mathbb{R}^n)$.*
- *$DH : \mathbb{R}^n \to \mathbb{R}^n$ is a C^{k-1} diffeomorphism.*
- *$DL(v) = (DH)^{-1}(v)$, $D^2L(v) = \left[D^2H(DL(v))\right]^{-1}$ for $v \in \mathbb{R}^n$.*
- *$L(v) = v \cdot DL(v) - H(DL(v))$ for $v \in \mathbb{R}^n$.*

Proof. As H is locally uniformly convex, it is strictly convex, and thus, $L \in C^1$ by Theorem 1.12. In the current setting, (1.3) becomes

$$p = DL(v) \qquad \Longleftrightarrow \qquad v = DH(p).$$

Thus, $(DL)^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ is well-defined and $(DL)^{-1} = DH$, which is of class C^{k-1}. Since $D^2H(p) > 0$ for all $p \in \mathbb{R}^n$, we use the inverse function theorem to deduce that $DH : \mathbb{R}^n \to \mathbb{R}^n$ is a local C^{k-1} diffeomorphism. Therefore, DL is also a local C^{k-1} diffeomorphism, and L is of class C^k. By definition, $DH(DL(v)) = v$ for all $v \in \mathbb{R}^n$, and hence,

$$D^2H(DL(v)) \cdot D^2L(v) = I_n.$$

Here, I_n is the identity matrix of size n. We conclude that

$$D^2L(v) = \left[D^2H(DL(v))\right]^{-1}.$$

$\square$

It is important to note that when $H, L \in C^k(\mathbb{R}^n)$ for some $k \geq 2$, the identity $DL = (DH)^{-1}$ shows a key property of the duality between H and L, and this also explains intuitively why $H^{**} = L^* = H$. In fact, in some literature, the Legendre transform was defined by this key property.

1.3. Hamiltonians depending on positions

We consider Hamiltonians that depend also on positions, that is, $H = H(x,p)$ for $(x,p) \in \mathbb{R}^n \times \mathbb{R}^n$. In this section, we always assume the following.

$$\begin{cases} H \in C(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R}), p \mapsto H(x,p) \text{ is convex for each } x \in \mathbb{R}^n, \\ \lim_{|p|\to\infty} \left(\inf_{x \in \mathbb{R}^n} \frac{H(x,p)}{|p|}\right) = +\infty. \end{cases} \tag{1.5}$$

The second condition in (1.5) is often called the uniform superlinearity of the Hamiltonian. We define the Lagrangian as

$$L(x,v) = H^*(x,v) = \sup_{p \in \mathbb{R}^n} \big(p \cdot v - H(x,p)\big) \quad \text{for } (x,v) \in \mathbb{R}^n \times \mathbb{R}^n. \tag{1.6}$$

Theorem 1.14. *Assume* (1.5). *Then, the following properties hold.*

(i) $L \in C(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R})$ *and* $v \mapsto L(x,v)$ *is convex and superlinear in* v *for each fixed* $x \in \mathbb{R}^n$.

(ii) $L^* = H^{**} = H$.

(iii) *For each* $R > 0$, *there exists* $C_R > 0$ *such that, for* $(x,v) \in \overline{B}(0,R) \times \overline{B}(0,R)$,

$$L(x,v) = \max_{|p| \leq C_R} \big(p \cdot v - H(x,p)\big).$$

(iv) *If* H *is strictly convex in* p, *then* $D_vL(x,v)$ *exists and* $(x,v) \mapsto D_vL(x,v)$ *is continuous.*

Proof. The only new thing to prove in (i) is the continuity of L as other points follow from Lemma 1.5. Assume $(x_k, v_k) \to (x_0, v_0)$ in $\mathbb{R}^n \times \mathbb{R}^n$.

There exists $C > 0$ such that $|x_k| + |v_k| \le C$ for all $k \in \mathbb{N}$. For each $k \in \mathbb{N}$, we are able to find $p_k \in \mathbb{R}^n$ with $|p_k| \le C$ such that

$$L(x_k, v_k) = p_k \cdot v_k - H(x_k, p_k).$$

For $k \in \mathbb{N}$, set

$$\omega(k) = |H(x_0, p_k) - H(x_k, p_k)| + C|v_k - v_0|.$$

It is clear that $\lim_{k\to\infty} \omega(k) = 0$. By Fenchel's inequality,

$$\begin{aligned} L(x_k, v_k) &= p_k \cdot v_k - H(x_k, p_k) \\ &\le p_k \cdot v_0 - H(x_0, p_k) + \omega(k) \le L(x_0, v_0) + \omega(k). \end{aligned}$$

Therefore

$$\limsup_{k\to\infty} L(x_k, v_k) \le L(x_0, v_0).$$

For each $p \in \mathbb{R}^n$ and $k \in \mathbb{N}$,

$$L(x_k, v_k) \ge p \cdot v_k - H(x_k, p),$$

which gives us that

$$\liminf_{k\to\infty} L(x_k, v_k) \ge p \cdot v_0 - H(x_0, p) \quad \Longrightarrow \quad \liminf_{k\to\infty} L(x_k, v_k) \ge L(x_0, v_0).$$

We obtain that $L \in C(\mathbb{R}^n \times \mathbb{R}^n, \mathbb{R})$.

For other parts, (ii) follows from Lemma 1.7, (iii) follows from Lemma 1.11, and (iv) is deduced from Theorem 1.12. $\square$

Theorem 1.15. *Assume* (1.5)*. Assume further that $H \in C^k(\mathbb{R}^n \times \mathbb{R}^n)$ for $k \ge 2$, and H is locally uniformly convex in p, i.e., $D^2_{pp}H(x, p) > 0$ for all $(x, p) \in \mathbb{R}^n \times \mathbb{R}^n$. Then, $L \in C^k(\mathbb{R}^n \times \mathbb{R}^n)$ and, for each $(x, v) \in \mathbb{R}^n \times \mathbb{R}^n$, there exists a unique $p(x, v) \in \mathbb{R}^n$ such that*

$$\begin{aligned} &p(x, v) = D_v L(x, v) \\ &D_x L(x, v) = -D_x H(x, p(x, v)) \\ &D^2_{vv} L(x, v) = \left[D^2_{pp} H(x, p(x, v))\right]^{-1}. \end{aligned}$$

Also, $p(x, v) = D_v L(x, v)$ implies $v = D_p H(x, p(x, v))$.

Proof. By Theorem 1.13, for each $x \in \mathbb{R}^n$, $v \mapsto L(x, v)$ is of class C^k. The relation

$$p = D_v L(x, v) \quad \Longleftrightarrow \quad v = D_p H(x, p)$$

defines a map $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$

$$\mathcal{L}(x, v) = (x, p) = (x, D_v L(x, v))$$

with its inverse $\mathcal{H} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$

$$\mathcal{H}(x, p) = (x, D_p H(x, p)).$$

By the given assumptions and the inverse function theorem, $\mathcal{H}$ is a $C^{k-1}(\mathbb{R}^n\times\mathbb{R}^n)$ diffeomorphism. Hence, $\mathcal{L}$ is also is a $C^{k-1}(\mathbb{R}^n\times\mathbb{R}^n)$ diffeomorphism, and in particular,

$$(x,v)\mapsto p(x,v)=D_vL(x,v)\in C^{k-1}(\mathbb{R}^n\times\mathbb{R}^n).$$

We need to show that $(x,v)\mapsto D_xL(x,v)$ is $C^{k-1}(\mathbb{R}^n\times\mathbb{R}^n)$. From the identity

$$L(x,v)=p(x,v)\cdot v-H(x,p(x,v))$$

we deduce that $x\mapsto L(x,v)$ is C^1 for each $v\in\mathbb{R}^n$. Differentiating this equality with respect to x to yield

$$\begin{aligned}D_xL(x,v)&=-D_xH(x,p(x,v))+v\cdot D_xp(x,v)-D_pH(x,p(x,v))\cdot D_xp(x,v)\\&=-D_xH(x,p(x,v))\end{aligned}$$

since $v=D_pH(x,p(x,v))$. As $D_xH,p\in C^{k-1}(\mathbb{R}^n\times\mathbb{R}^n)$, we get $D_xL\in C^{k-1}(\mathbb{R}^n\times\mathbb{R}^n)$. Hence, $L\in C^k(\mathbb{R}^n\times\mathbb{R}^n)$. □

Definition 1.16. Define

$$\begin{aligned}\mathcal{H}:\mathbb{R}^n\times\mathbb{R}^n&\to\mathbb{R}^n\times\mathbb{R}^n\\(x,p)&\mapsto(x,v)=(x,D_pH(x,p)),\end{aligned}$$

and its inverse (dual)

$$\begin{aligned}\mathcal{L}:\mathbb{R}^n\times\mathbb{R}^n&\to\mathbb{R}^n\times\mathbb{R}^n\\(x,v)&\mapsto(x,p)=(x,D_vL(x,v)).\end{aligned}$$

Under the assumptions of Theorem 1.15, $\mathcal{H},\mathcal{L}$ are both local C^{k-1} diffeomorphisms.

Remark 1.17. Sometimes, we assume more that L is bounded in $\mathbb{R}^n\times\overline{B}(0,R)$ for each $R>0$ to get the boundedness of $p(x,v)\in\partial_vL(x,v)$ for $(x,v)\in\mathbb{R}^n\times\overline{B}(0,R)$. Indeed, as $p(x,v)\in\partial_vL(x,v)$,

$$L(x,v+h)\ge L(x,v)+p(x,v)\cdot h\qquad\text{for all } h\in\mathbb{R}^n.$$

In particular,

$$|p(x,v)|=\max_{|h|\le1}p(x,v)\cdot h\le|L(x,v)|+|L(x,v+h)|\le2\sup_{\mathbb{R}^n\times\overline{B}(0,R+1)}L.$$

Let us state this as a theorem.

Theorem 1.18. *Assume* (1.5) *and* $H\in L^\infty\left(\mathbb{R}^n\times\overline{B}(0,R)\right)$ *for each* $R>0$. *Then, the following properties hold.*

(i) $L\in L^\infty\left(\mathbb{R}^n\times\overline{B}(0,R)\right)$ *for each* $R>0$, *and* L *satisfies*

$$\begin{cases}L\in C(\mathbb{R}^n\times\mathbb{R}^n,\mathbb{R}),v\mapsto L(x,v)\ \textit{is convex for each}\ x\in\mathbb{R}^n,\\\lim_{|v|\to\infty}\left(\inf_{x\in\mathbb{R}^n}\frac{L(x,v)}{|v|}\right)=+\infty.\end{cases}$$

(ii) *For each $R > 0$, there exists $C_R > 0$ such that, for $(x,v) \in \mathbb{R}^n \times \overline{B}(0,R)$,*

$$L(x,v) = \max_{|p| \le C_R} \big(p \cdot v - H(x,p)\big).$$

Proof. For $R > 0$, denote by

$$C_H(R) = \|H\|_{L^\infty\left(\mathbb{R}^n \times \overline{B}(0,R)\right)}.$$

We first show that $L \in L^\infty\left(\mathbb{R}^n \times \overline{B}(0,R)\right)$. Fix $(x,v) \in \mathbb{R}^n \times \overline{B}(0,R)$. By definition of the Legendre transform,

$$L(x,v) = \sup_{p \in \mathbb{R}^n} (p \cdot v - H(x,p)) \ge -H(x,0) \ge -C_H(1).$$

Thanks to the uniform superlinearity of H in (1.5), there exists $C = C(H,R) > 1$ such that

$$\inf_{x \in \mathbb{R}^n} \frac{H(x,p)}{|p|} > R + C_H(1) \quad \text{for } |p| \ge C,$$

which implies

$$H(x,p) > R|p| + C_H(1) \quad \text{for } |p| \ge C.$$

Therefore,

$$\begin{aligned} L(x,v) &= \sup_{p \in \mathbb{R}^n} (p \cdot v - H(x,p)) \le \sup_{p \in \mathbb{R}^n} (R|p| - H(x,p)) \\ &= \sup_{|p| \le C} (R|p| - H(x,p)) \le CR + C_H(C). \end{aligned}$$

Thus, $L \in L^\infty\left(\mathbb{R}^n \times \overline{B}(0,R)\right)$.

Next, we show that L is uniformly superlinear in v. For $R > 0$ and $v \ne 0$, by choosing $p = R\frac{v}{|v|}$, we see that

$$L(x,v) = \sup_{p \in \mathbb{R}^n} (p \cdot v - H(x,p)) \ge R|v| - C_H(R).$$

Hence,

$$\liminf_{|v| \to \infty} \left(\inf_{x \in \mathbb{R}^n} \frac{L(x,v)}{|v|} \right) \ge \liminf_{|v| \to \infty} \frac{R|v| - C_H(R)}{|v|} \ge R.$$

We let $R \to \infty$ to confirm that L is uniformly superlinear in v.

Claim (ii) follows from Remark 1.17. □

1.4. The Legendre transform and other transformations

We first study the Legendre transform under scalings and translations.

Lemma 1.19. *Let $H : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Let $a > 0$ and $b \in \mathbb{R}$ be given numbers. Denote by $G = aH + b$. Then, for $v \in \mathbb{R}^n$,*

$$G^*(v) = aH^*\left(\frac{v}{a}\right) - b.$$

Proof. We compute, for $v \in \mathbb{R}^n$,

$$\begin{aligned} G^*(v) &= \sup_{p\in\mathbb{R}^n} (p \cdot v - G(p)) \\ &= \sup_{p\in\mathbb{R}^n} (p \cdot v - aH(p) - b) \\ &= a \sup_{p\in\mathbb{R}^n} \left(p \cdot \frac{v}{a} - H(p)\right) - b = aH^*\left(\frac{v}{a}\right) - b. \end{aligned}$$

$\square$

Lemma 1.20. *Let $H : \mathbb{R}^n \to \mathbb{R}$ be a convex function. Let A be a symmetric, invertible $n \times n$ matrix. Denote by $G(p) = H(Ap)$ for $p \in \mathbb{R}^n$. Then, for $v \in \mathbb{R}^n$,*

$$G^*(v) = H^*\left(A^{-1}v\right).$$

Proof. For $v \in \mathbb{R}^n$,

$$\begin{aligned} G^*(v) &= \sup_{p\in\mathbb{R}^n} (p \cdot v - G(p)) \\ &= \sup_{p\in\mathbb{R}^n} \left(Ap \cdot A^{-1}v - H(Ap)\right) = H^*\left(A^{-1}v\right). \end{aligned}$$

$\square$

Next, we consider the Legendre transform under infimal convolutions.

Definition 1.21 (Infimal convolutions)**.** Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be two given functions. The infimal convolution $f *_{\text{inf}} g$ is defined as, for $x \in \mathbb{R}^n$,

$$(f *_{\text{inf}} g)(x) = \inf\{f(x-y) + g(y) \,:\, y \in \mathbb{R}^n\}.$$

Proposition 1.22. *Let $H, G : \mathbb{R}^n \to \mathbb{R}$ be convex functions. Then, $H *_{\text{inf}} G$ is convex, and*

$$(H *_{\text{inf}} G)^* = H^* + G^*.$$

Proof. We first show that $H *_{\text{inf}} G$ is convex. Indeed, for $p_1, p_2, q_1, q_2 \in \mathbb{R}^n$,

$$\begin{aligned} &H(p_1 - q_1) + G(q_1) + H(p_2 - q_2) + G(q_2) \\ &\geq 2\left(H\left(\frac{p_1+p_2}{2} - \frac{q_1+q_2}{2}\right) + G\left(\frac{q_1+q_2}{2}\right)\right) \geq 2(H *_{\text{inf}} G)\left(\frac{p_1+p_2}{2}\right). \end{aligned}$$

Take infimum over $q_1, q_2 \in \mathbb{R}^n$ in the above to yield

$$(H *_{\text{inf}} G)(p_1) + (H *_{\text{inf}} G)(p_2) \geq 2(H *_{\text{inf}} G)\left(\frac{p_1+p_2}{2}\right),$$

and hence, $H *_{\text{inf}} G$ is convex.

Next, we compute, for $v \in \mathbb{R}^n$,

$$\begin{aligned}(H *_{\text{inf}} G)^*(v) &= \sup_{p\in\mathbb{R}^n}\left(p\cdot v - \inf_{q\in\mathbb{R}^n}(H(p-q)+G(q))\right)\\ &= \sup_{p,q\in\mathbb{R}^n}((p-q)\cdot v - H(p-q) + q\cdot v - G(q))\\ &= H^*(v)+G^*(v).\end{aligned}$$

□

1.5. References

(1) The content on the Legendre transform covered in this chapter is quite classical. We also refer the readers to other books for similar material [**CS04, Eva10, Tra21**].

(2) For a characterization of the Legendre transform, see Artstein-Avidan and Milman [**AAM09**], which is covered in [**Tra21**, Appendix C]. Basically, up to translations and a linear change of variables, the Legendre transform is the unique transform that is involutive and reverses the order in the set of convex functions.

Action functionals and their minimizers

In this chapter, we always consider a given Lagrangian $L : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ that satisfies

$$\begin{cases} L \in C^k(\mathbb{R}^n \times \mathbb{R}^n) \text{ for some } k \geq 2, \\ D^2_{vv}L(x,v) > 0 \text{ for all } (x,v) \in \mathbb{R}^n \times \mathbb{R}^n, \\ \lim_{|v|\to\infty} \inf_{x\in\mathbb{R}^n} \frac{L(x,v)}{|v|} = +\infty. \end{cases} \tag{2.1}$$

From Theorem 1.15 and Definition 1.16, we have that

$$\begin{aligned} \mathcal{L} : \mathbb{R}^n \times \mathbb{R}^n &\to \mathbb{R}^n \times \mathbb{R}^n \\ (x,v) &\mapsto (x,p) = (x, D_vL(x,v)). \end{aligned} \tag{2.2}$$

is a local C^{k-1} diffeomorphism thanks to (2.1). The inverse of $\mathcal{L}$ is $\mathcal{H}$,

$$\begin{aligned} \mathcal{H} : \mathbb{R}^n \times \mathbb{R}^n &\to \mathbb{R}^n \times \mathbb{R}^n \\ (x,p) &\mapsto (x,v) = (x, D_pH(x,p)), \end{aligned}$$

which is also a local C^{k-1} diffeomorphism. Here, H is the Legendre transform of L.

Our main object here is the action functional

$$I[\gamma] = \int_a^b L(\gamma(t), \dot{\gamma}(t))\, dt,$$

where $a < b$ are two given real numbers, and $\gamma : [a,b] \to \mathbb{R}^n$ is a given curve belonging to certain admissible class $\mathcal{A}$ to be specified. The problem of interests is the following calculus of variation problem

$$\min_{\gamma\in\mathcal{A}} I[\gamma] = \min_{\gamma\in\mathcal{A}} \int_a^b L(\gamma(t),\dot\gamma(t))\,dt.$$

There are various different ways to choose the admissible class $\mathcal{A}$ and we will start with the most basic/classical one.

2.1. Minimizers and the Euler-Lagrange equations in the class of continuous and piecewise C^1 curves

We first give a definition of admissible class of continuous and piecewise C^1 curves.

Definition 2.1 (Admissible class of continuous and piecewise C^1 curves)**.** Let $a<b$ be two given real numbers, and $y,z\in\mathbb{R}^n$ be two given vectors. Denote by

$$\mathcal{A} = \big\{\gamma : [a,b]\to\mathbb{R}^n \,:\, \gamma(a)=y, \gamma(b)=z,\\ \gamma \text{ is a continuous, piecewise } C^1 \text{ curve}\big\}.$$

Definition 2.2 (Action of a curve)**.** Let $\gamma\in\mathcal{A}$. Then, the action of γ for L is

$$I[\gamma] = \int_a^b L(\gamma(t),\dot\gamma(t))\,dt.$$

Definition 2.3 (Minimizers in $\mathcal{A}$)**.** We say that $\gamma\in\mathcal{A}$ is a minimizer of the action for L in the admissible class $\mathcal{A}$ if

$$I[\gamma] = \min_{\eta\in\mathcal{A}} I[\eta]. \tag{2.3}$$

2.1.1. Euler-Lagrange equations for minimizers.

Theorem 2.4. *Assume* (2.1)*. Let $\gamma\in\mathcal{A}\cap C^2([a,b],\mathbb{R}^n)$ be a minimizer of the action for L in the admissible class $\mathcal{A}$. Then, γ satisfies*

$$\frac{d}{dt}\left(D_vL(\gamma(t),\dot\gamma(t))\right) = D_xL(\gamma(t),\dot\gamma(t)) \quad \textit{for } a\le t\le b. \tag{2.4}$$

Proof. The proof is quite classical via the variational method. Fix $\eta : [a,b]\to\mathbb{R}^n$ smooth such that $\eta(a)=\eta(b)=0$. Then, for $s\in\mathbb{R}^n$, $\gamma+s\eta\in\mathcal{A}$. Define $i:\mathbb{R}\to\mathbb{R}$ as

$$i(s) = I[\gamma+s\eta].$$

By a straightforward computation,

$$i'(s) = \int_a^b \left(D_xL(\gamma+s\eta,\dot\gamma+s\dot\eta)\cdot\eta + D_vL(\gamma+s\eta,\dot\gamma+s\dot\eta)\cdot\dot\eta\right)\,dt$$

Thanks to (6.5), $i(0) = \min_{s\in\mathbb{R}} i(s)$. In particular,

$$\begin{aligned} 0 = i'(0) &= \int_a^b \left(D_x L(\gamma,\dot\gamma)\cdot\eta + D_v L(\gamma,\dot\gamma)\cdot\dot\eta\right) dt \\ &= \int_a^b \left(D_x L(\gamma,\dot\gamma) - \frac{d}{dt}\left(D_v L(\gamma(t),\dot\gamma(t))\right)\right)\cdot\eta\, dt, \end{aligned}$$

where we used integration by parts and $\eta(a) = \eta(b) = 0$ in the last equality. Since the above holds true for all smooth $\eta : [a,b] \to \mathbb{R}^n$ with $\eta(a) = \eta(b) = 0$, we imply that (2.4) holds.

$\square$

Remark 2.5. As γ is C^2, we are able to expand the Euler-Lagrange equations (2.4) out as

$$D^2_{xv} L(\gamma,\dot\gamma)\dot\gamma + D^2_{vv} L(\gamma,\dot\gamma)\ddot\gamma = D_x L(\gamma,\dot\gamma),$$

which is equivalent to

$$\ddot\gamma = \left(D^2_{vv} L(\gamma,\dot\gamma)\right)^{-1}\left(D_x L(\gamma,\dot\gamma) - D^2_{xv} L(\gamma,\dot\gamma)\dot\gamma\right) \quad \text{on } [a,b].$$

It is clear that we employ the C^2 assumption of γ strongly in the proof of Theorem 2.4 above to have the term $\frac{d}{dt}\left(D_v L(\gamma(t),\dot\gamma(t))\right)$ defined in the classical way to have the Euler-Lagrange equations. We now proceed to show that we can relax this C^2 regularity requirement, and requiring C^1 regularity is enough.

Theorem 2.6. *Assume* (2.1). *Let $\gamma \in \mathcal{A} \cap C^1([a,b],\mathbb{R}^n)$ be a minimizer of the action for L in the admissible class $\mathcal{A}$. Then, $\gamma \in C^2$, and in fact, $\gamma \in C^k$.*

Proof. Following the proof of Theorem 2.4, we have

$$0 = i'(0) = \int_a^b \left(D_x L(\gamma,\dot\gamma)\cdot\eta + D_v L(\gamma,\dot\gamma)\cdot\dot\eta\right) dt.$$

As we only have $\gamma \in C^1$, we integrate by parts in a different way to get

$$\int_a^b \left(-\int_a^t D_x L(\gamma,\dot\gamma)\, ds + D_v L(\gamma,\dot\gamma)\right)\cdot\dot\eta\, dt = 0.$$

As the above holds true for all $\eta : [a,b] \to \mathbb{R}^n$ smooth with $\eta(a) = \eta(b) = 0$, there exists a vector $q \in \mathbb{R}^n$ such that

$$D_v L(\gamma(t),\dot\gamma(t)) = q + \int_a^t D_x L(\gamma(s),\dot\gamma(s))\, ds \quad \text{for } t \in [a,b]. \tag{2.5}$$

As $\gamma \in C^1$, we use (2.5) to deduce that

$$t \mapsto D_v L(\gamma(t),\dot\gamma(t)) \quad \text{is in } C^1. \tag{2.6}$$

Recall that $\mathcal{L}$ is a local C^{k-1} diffeomorphism by (2.2), and $\mathcal{H} \in C^{k-1}$ is its inverse, and

$$\mathcal{H}\left(\gamma(t), D_v L(\gamma(t), \dot{\gamma}(t))\right) = (\gamma(t), \dot{\gamma}(t)) \quad \text{for } t \in [a,b]. \tag{2.7}$$

We combine (2.6) and (2.7) to yield $t \mapsto (\gamma(t), \dot{\gamma}(t))$ is C^1, which means that $\gamma \in C^2$.

By induction, we deduce further that $t \mapsto (\gamma(t), \dot{\gamma}(t))$ is C^{k-1}, which implies that $\gamma \in C^k$. The proof is complete. $\square$

We next show that the C^1 regularity assumption can be weakened to the piecewise C^1 regularity assumption, which is exactly the regularity condition we put on our admissible class $\mathcal{A}$.

Theorem 2.7. *Assume* (2.1). *Let $\gamma \in \mathcal{A}$ be a minimizer of the action for L in the admissible class $\mathcal{A}$. Then, $\gamma \in C^k$.*

Proof. We can find $a = a_0 < a_1 < \ldots < a_m = b$ such that γ is C^1 on $[a_i, a_{i+1}]$ for $0 \le i \le m-1$. By the assumption, γ is a minimizer of the action for L on $[a_i, a_{i+1}]$ for $0 \le i \le m-1$. Thanks to Theorem 2.6, $\gamma \in C^k([a_i, a_{i+1}])$ for $0 \le i \le m-1$. Moreover, γ satisfies the Euler-Lagrange equations (2.4) on $[a_i, a_{i+1}]$ for $0 \le i \le m-1$.

By the calculus of variations method in the proof of Theorem 2.4, we have, for all $\eta : [a,b] \to \mathbb{R}^n$ smooth with $\eta(a) = \eta(b) = 0$,

$$0 = i'(0) = \int_a^b \left(D_x L(\gamma, \dot{\gamma}) \cdot \eta + D_v L(\gamma, \dot{\gamma}) \cdot \dot{\eta}\right) dt.$$

It is rather natural to split the above integral on $[a,b]$ to integrals on $[a_i, a_{i+1}]$ for $0 \le i \le m-1$. Indeed,

$$\begin{aligned}
0 &= \int_a^b \left(D_x L(\gamma, \dot{\gamma}) \cdot \eta + D_v L(\gamma, \dot{\gamma}) \cdot \dot{\eta}\right) dt \\
&= \sum_{i=0}^{m-1} \int_{a_i}^{a_{i+1}} \left(D_x L(\gamma, \dot{\gamma}) \cdot \eta + D_v L(\gamma, \dot{\gamma}) \cdot \dot{\eta}\right) dt \\
&= \sum_{i=0}^{m-1} \int_{a_i}^{a_{i+1}} \left(D_x L(\gamma, \dot{\gamma}) - \frac{d}{dt}\left(D_v L(\gamma, \dot{\gamma})\right)\right) \cdot \eta \, dt \\
&\qquad + \sum_{i=0}^{m-1} \left(D_v L(\gamma(a_{i+1}), \dot{\gamma}(a_{i+1}^-)) \cdot \eta(a_{i+1}) - D_v L(\gamma(a_i), \dot{\gamma}(a_i^+)) \cdot \eta(a_i)\right) \\
&= \sum_{i=1}^{m-1} \left(D_v L(\gamma(a_i), \dot{\gamma}(a_i^-)) - D_v L(\gamma(a_i), \dot{\gamma}(a_i^+))\right) \cdot \eta(a_i).
\end{aligned}$$

We used the fact that γ satisfies the Euler-Lagrange equations (2.4) on $[a_i, a_{i+1}]$ for $0 \le i \le m-1$, and $\eta(a) = \eta(b) = 0$ in the last equality above. As $\eta(a_i)$ can be chosen arbitrarily for $1 \le i \le m-1$, we yield

$$D_v L(\gamma(a_i), \dot\gamma(a_i^-)) = D_v L(\gamma(a_i), \dot\gamma(a_i^+)),$$

which gives that $\dot\gamma(a_i^-) = \dot\gamma(a_i^+)$. Thus, $\gamma \in C^1([a,b])$. Applying Theorem 2.6 once more, we conclude that $\gamma \in C^k([a,b])$. □

2.1.2. Extremal curves.

Definition 2.8 (Extremal curves in $\mathcal{A}$)**.** We say that $\gamma \in \mathcal{A}$ is an extremal curve of the action for L in the admissible class $\mathcal{A}$ if we have, for all smooth $\eta : [a,b] \to \mathbb{R}^n$ with $\eta(a) = \eta(b) = 0$,

$$\frac{d}{ds} I[\gamma + s\eta]\Big|_{s=0} = 0. \tag{2.8}$$

It is clear that Theorem 2.7 holds for extremal curves as well. We state this result here for clarity and for usage later.

Theorem 2.9. *Assume* (2.1)*. Let $\gamma \in \mathcal{A}$ be an extremal curve of the action for L in the admissible class $\mathcal{A}$. Then, $\gamma \in C^k$.*

2.2. Connections between Lagrangian and Hamiltonian viewpoints

In this section, we always assume (2.1). We have shown in Theorems 2.7 and 2.9 that if γ is an extremal curve or a minimizer of the action for L and γ is continuous and piecewise C^1, then $\gamma \in C^k([a,b])$, and γ satisfies the Euler-Lagrange equations

$$\frac{d}{dt}\left(D_v L(\gamma(t), \dot\gamma(t))\right) = D_x L(\gamma(t), \dot\gamma(t)) \quad \text{for } a \le t \le b. \tag{2.9}$$

2.2.1. Hamiltonian system. We provide connection between this viewpoint and the Hamiltonian dynamics. Denote by, for $t \in [a,b]$,

$$\begin{cases} X(t) = \gamma(t), \\ P(t) = D_v L(\gamma(t), \dot\gamma(t)). \end{cases} \tag{2.10}$$

By the Legendre transform,

$$L(\gamma(t), \dot\gamma(t)) + H(\gamma(t), D_v L(\gamma(t), \dot\gamma(t))) = \dot\gamma(t) \cdot D_v L(\gamma(t), \dot\gamma(t)),$$

which means that

$$\begin{cases} L(\gamma(t), \dot\gamma(t)) + H(X(t), P(t)) = \dot\gamma(t) \cdot P(t), \\ \dot\gamma(t) = D_p H(X(t), P(t)). \end{cases}$$

We use the above and (2.9) to yield, for $t \in [a,b]$,

$$\begin{cases} \dot{X}(t) = \dot{\gamma}(t) = D_pH(X(t),P(t)), \\ \dot{P}(t) = \frac{d}{dt}\left(D_vL(\gamma(t),\dot{\gamma}(t))\right) = D_xL(\gamma(t),\dot{\gamma}(t)) = -D_xH(X(t),P(t)). \end{cases}$$

Therefore, $(X(t),P(t))$ satisfies the following Hamiltonian system, for $t \in [a,b]$,

$$\begin{cases} \dot{X}(t) = D_pH(X(t),P(t)), \\ \dot{P}(t) = -D_xH(X(t),P(t)). \end{cases} \tag{2.11}$$

Remark 2.10. In terms of our notations on dynamics of a given particle in Lagrangian and Hamiltonian viewpoints, we have the following.

- We typically use $\gamma(t)$ in Lagrangian framework and $X(t)$ in Hamiltonian framework to represent the position of this particle at time $t \in \mathbb{R}$. The variable x stands for the position variable.
- The corresponding velocity at time t is $\dot{\gamma}(t)$ in Lagrangian framework. And the velocity variable is v.
- The generalized momentum of this particle at time t is $P(t)$ with the relation $P(t) = D_vL(\gamma(t),\dot{\gamma}(t))$. The variable p represents for the momentum variable correspondingly.

Let us now consider the classical mechanics setting where we normalize this particle to be of unit mass, that is, $m = 1$. Then,

$$H(x,p) = \frac{1}{2}|p|^2 + V(x) \quad \text{and} \quad L(x,v) = \frac{1}{2}|v|^2 - V(x),$$

where V is the potential energy. In this situation,

$$P(t) = D_vL(\gamma(t),\dot{\gamma}(t)) = \dot{\gamma}(t) = m\dot{\gamma}(t),$$

which is precisely the classical momentum of the particle at time t.

Lemma 2.11. *Let $(X(t),P(t))$ be a solution to (2.11) for $t \in [a,b]$. Then, $t \mapsto H(X(t),P(t))$ is constant on $[a,b]$.*

Proof. We calculate that

$$\begin{aligned} &\frac{d}{dt}\left(H(X(t),P(t))\right) = D_xH(X(t),P(t)) \cdot \dot{X}(t) + D_pH(X(t),P(t)) \cdot \dot{P}(t) \\ = &\ \left[D_xH \cdot D_pH + D_pH \cdot (-D_xH)\right](X(t),P(t)) = 0. \end{aligned}$$

Hence, $t \mapsto H(X(t),P(t))$ is constant on $[a,b]$.

$\square$

Remark 2.12. The Hamiltonian H stands for total energy, and it is rather natural that the total energy is conserved along the Hamiltonian flow. For given $(X(a), P(a))$, we then see that

$$H(X(t), P(t)) = H(X(a), P(a)) \leq C \quad \text{for all } t \in [a, b].$$

By the superlinearity of H in p, we yield

$$|P(t)| \leq C \quad \text{for all } t \in [a, b].$$

Recall that $L(\gamma(t), \dot{\gamma}(t)) + H(X(t), P(t)) = \dot{\gamma}(t) \cdot P(t)$, which together with the inequality above implies

$$L(\gamma(t), \dot{\gamma}(t)) \leq C + C|\dot{\gamma}(t)| \quad \text{for all } t \in [a, b].$$

We employ the superlinearity of L in v to deduce

$$|\dot{\gamma}(t)| \leq C \quad \Longrightarrow \quad |\gamma(t)| \leq C \quad \text{for all } t \in [a, b].$$

Thus,

$$|X(t)| + |P(t)| \leq C \quad \text{for all } t \in [a, b]. \tag{2.12}$$

As $D_x H, D_p H \in \mathrm{Lip}\,(\overline{B}(0, C) \times \overline{B}(0, C))$, we see that the Hamiltonian system (2.11) is wellposed.

2.2.2. Poisson's bracket.

Definition 2.13 (Poisson's bracket)**.** For $f, g \in C^1(\mathbb{R}^n \times \mathbb{R}^n)$, we define the Poisson bracket $\{f, g\}$ as

$$\{f, g\}(x, p) = D_x f(x, p) \cdot D_p g(x, p) - D_p f(x, p) \cdot D_x g(x, p).$$

We record in the following some basic properties of the Poisson bracket.

Proposition 2.14 (Basic properties of the Poisson bracket)**.** *Let $f, g, h \in C^2(\mathbb{R}^n \times \mathbb{R}^n)$ and $a, b \in \mathbb{R}$. The following properties hold.*

(i) *Anticommutativity*

$$\{f, g\} = -\{g, f\}.$$

(ii) *Bilinearity*

$$\{af + bg, h\} = a\{f, h\} + b\{g, h\}.$$

(iii) *Leibniz's rule*

$$\{fg, h\} = f\{g, h\} + g\{f, h\}.$$

(iv) *Jacobi's identity*

$$\{f, \{g, h\}\} + \{g, \{h, f\}\} + \{h, \{f, g\}\} = 0.$$

We have further the following lemma.

Lemma 2.15. *Let $\phi \in C^1(\mathbb{R}^n \times \mathbb{R}^n)$. Let $(X(t), P(t))$ be a solution to* (2.11) *for $t \in [a, b]$. Then,*

$$\frac{d}{dt}\left(\phi(X(t), P(t))\right) = \{\phi, H\}(X(t), P(t)).$$

In particular, if $\{\phi, H\} = 0$, then $t \mapsto \phi(X(t), P(t))$ is constant on $[a, b]$.

Proof. We compute

$$\begin{aligned} &\frac{d}{dt}\left(\phi(X(t), P(t))\right) = D_x\phi(X(t), P(t)) \cdot \dot{X}(t) + D_p\phi(X(t), P(t)) \cdot \dot{P}(t) \\ = \;& \left[D_x\phi \cdot D_pH + D_p\phi \cdot (-D_xH)\right](X(t), P(t)) = \{\phi, H\}(X(t), P(t)). \end{aligned}$$

The proof is complete. □

2.2.3. Lagrangian and Hamiltonian flows.

Definition 2.16 (Lagrangian and Hamiltonian flows)**.** The Lagrangian flow $\phi_t^L : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ for $t \in \mathbb{R}$ is defined as

$$\begin{cases} \phi_t^L(x, v) = (\gamma(t), \dot{\gamma}(t)), \\ (\gamma(0), \dot{\gamma}(0)) = (x, v), \end{cases}$$

where γ satisfies the Euler-Lagrange equations (2.9) in $\mathbb{R}$.

The Hamiltonian flow $\phi_t^H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ for $t \in \mathbb{R}$ is defined as

$$\begin{cases} \phi_t^L(x, p) = (X(t), P(t)), \\ (X(0), P(0)) = (x, p), \end{cases}$$

where (X, P) solves the Hamiltonian system (2.11) in $\mathbb{R}$.

By using the local C^{k-1} diffeomorphism $(x, v) \mapsto \mathcal{L}(x, v) = (x, p) = (x, D_vL(x, v))$, we have an important identity that

$$\mathcal{L} \circ \phi_t^L \circ \mathcal{L}^{-1} = \phi_t^H. \tag{2.13}$$

This identity allows us to go back and forth between the Lagrangian flow and the Hamiltonian flow naturally.

2.3. Minimizers in the class of absolutely continuous curves

The space of continuous and piecewise C^1 curves with fixed endpoints is not compact (under a reasonable topology), and thus, it is more convenient to consider a bigger admissible class of curves, in which we have compactness. This is important for us to construct minimizers via a direct method in calculus of variations later.

2.3.1. Absolutely continuous curves.

Definition 2.17 (Absolutely continuous curves)**.** Let $\gamma \in C([a,b],\mathbb{R}^n)$. We say that γ is *absolutely continuous* if for each $\varepsilon > 0$ there exists $\delta > 0$ such that, if $\{(a_i,b_i)\}_{i\in\mathbb{N}}$ is a disjoint family of intervals in (a,b), then

$$\sum_{i\in\mathbb{N}} |b_i - a_i| < \delta \quad \Longrightarrow \quad \sum_{i\in\mathbb{N}} |\gamma(b_i) - \gamma(a_i)| < \varepsilon.$$

We have the following theorem on characterization of absolutely continuous curves, which is quite standard.

Theorem 2.18 (Characterization of absolutely continuous curves)**.** *Let $\gamma \in C([a,b],\mathbb{R}^n)$. Then, γ is absolutely continuous if and only if all of the following hold*

(i) *$\dot\gamma$ exists a.e. in (a,b);*

(ii) *$\dot\gamma$ is Lebesgue integrable on (a,b);*

(iii) *$\gamma(t) - \gamma(a) = \int_a^t \dot\gamma(s)\,ds$ for each $t \in [a,b]$.*

Definition 2.19 (Spaces of absolutely continuous curves)**.** Let $a < b$ be two given real numbers. Denote by

$$\mathrm{AC}([a,b],\mathbb{R}^n) = \{\gamma \in C([a,b],\mathbb{R}^n) \,:\, \gamma \text{ is absolutely continuous}\}.$$

The space $\mathrm{AC}([a,b],\mathbb{R}^n)$ of absolutely continuous curves enjoy the following compactness (*tightness*) property.

Theorem 2.20. *Let $\{\gamma_k\}_{k\in\mathbb{N}} \subset \mathrm{AC}([a,b],\mathbb{R}^n)$. Suppose that $\{\dot\gamma_k\}_{k\in\mathbb{N}}$ is uniformly integrable on $[a,b]$, that is for each $\varepsilon > 0$, there is $\delta > 0$ such that if $E \subset [a,b]$ is a Borel measurable set with measure $|E| < \delta$, then*

$$\sup_{k\in\mathbb{N}} \int_E |\dot\gamma_k(s)|\,ds < \varepsilon. \tag{2.14}$$

If there exists $t_0 \in [a,b]$ such that $\{\gamma_k(t_0)\}$ is bounded, then there exist a subsequence $\{\gamma_{k_j}\}$ of $\{\gamma_k\}$, and $\gamma \in \mathrm{AC}([a,b],\mathbb{R}^n)$ such that $\gamma_{k_j} \to \gamma$ uniformly on $[a,b]$ and $\dot\gamma_{k_j} \rightharpoonup \dot\gamma$ weakly in $L^1([a,b])$, that is

$$\lim_{j\to\infty} \int_a^b \dot\gamma_{k_j}(s)\cdot\phi(s)\,ds = \int_a^b \dot\gamma(s)\cdot\phi(s)\,ds$$

for all $\phi \in L^\infty([a,b],\mathbb{R}^n)$.

Proof. We split the proof into several steps for clarity.

Step 1. We first show that $\{\gamma_k\}$ is bounded and equi-continuous on $[a,b]$. By the hypothesis, for each $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$ such that, for

$|t_1 - t_2| < \delta(\varepsilon)$, we have

$$|\gamma_k(t_1) - \gamma_k(t_2)| \le \left| \int_{t_1}^{t_2} \dot{\gamma}_k(s)\, ds \right| < \varepsilon \tag{2.15}$$

for all $k \in \mathbb{N}$. Thus, $\{\gamma_k\}$ is equi-continuous on $[a, b]$.

On the other hand, as $\{\gamma_k(t_0)\}$ is bounded, there exists $C > 0$ such that $|\gamma_k(t_0)| \le C$ for all $k \in \mathbb{N}$. By using (2.15) repeatedly, we imply that, for all $k \in \mathbb{N}$ and $t \in [a, b]$,

$$|\gamma_k(t)| \le |\gamma_k(t_0)| + \left(\frac{b-a}{\delta(1)} + 1 \right) \quad \Longrightarrow \quad \|\gamma_k\|_{L^\infty([a,b])} \le C.$$

By the Arzelà-Ascoli theorem, there exists a subsequence $\{\gamma_{k_j}\}$ of $\{\gamma_k\}$ such that $\gamma_{k_j} \to \gamma$ uniformly on $[a, b]$. By abusing of notations, we write $\gamma_k \to \gamma$ uniformly on $[a, b]$.

Step 2. We now prove that γ is absolutely continuous. Fix $\varepsilon > 0$. Let $\{(a_i, b_i)\}_{i \in \mathbb{N}}$ be a sequence of disjoint open intervals with $\sum_{i \in \mathbb{N}} (b_i - a_i) < \delta(\varepsilon)$. Then, the tightness condition gives us that, for all $k \in \mathbb{N}$,

$$\sum_{i \in \mathbb{N}} |\gamma_k(b_i) - \gamma_k(a_i)| \le \sum_{i \in \mathbb{N}} \int_{a_i}^{b_i} |\dot{\gamma}_k(s)|\, ds < \varepsilon.$$

Let $k \to \infty$ to deduce that $\gamma \in \mathrm{AC}([a, b], \mathbb{R}^n)$.

Step 3. Finally, we show $\dot{\gamma}_k \rightharpoonup \dot{\gamma}$ weakly in $L^1([a, b])$. To obtain that

$$\lim_{k \to \infty} \int_a^b \dot{\gamma}_k(s) \phi(s)\, ds = \int_a^b \dot{\gamma}(s) \phi(s)\, ds \tag{2.16}$$

for $\phi \in L^\infty([a, b], \mathbb{R}^n)$, we approximate ϕ by simple functions from $[a, b]$ to $\mathbb{R}^n$. First of all, any open set U in (a, b) can be written as a disjoint union of countably many open sub-intervals $\{(a_i, b_i)\}_{i \in \mathbb{N}}$. For $\varepsilon > 0$, choose $m \in \mathbb{N}$ large enough such that $E = U \setminus \bigcup_{i=1}^m (a_i, b_i)$ has $|E| < \delta(\varepsilon)$. Then,

$$\sup_{k \in \mathbb{N}} \left| \int_U \dot{\gamma}_k(s)\, ds - \sum_{i=1}^m \int_{a_i}^{b_i} \dot{\gamma}_k(s)\, ds \right| < \varepsilon. \tag{2.17}$$

Besides,

$$\begin{aligned} \lim_{k \to \infty} \sum_{i=1}^m \int_{a_i}^{b_i} \dot{\gamma}_k(s)\, ds &= \lim_{k \to \infty} \sum_{i=1}^m \big(\gamma_k(b_i) - \gamma_k(a_i) \big) \\ &= \sum_{i=1}^m \big(\gamma(b_i) - \gamma(a_i) \big) = \sum_{i=1}^m \int_{a_i}^{b_i} \dot{\gamma}(s)\, ds. \end{aligned}$$

Using this and taking the limit as $k \to \infty$ in (2.17) to obtain

$$\sum_{i=1}^{m} \int_{a_i}^{b_i} \dot\gamma(s)\, ds - \varepsilon \le \liminf_{k\to\infty} \int_U \dot\gamma_k(s)\, ds$$
$$\le \limsup_{k\to\infty} \int_U \dot\gamma_k(s)\, ds \le \sum_{i=1}^{m} \int_{a_i}^{b_i} \dot\gamma(s)\, ds + \varepsilon.$$

Taking $\varepsilon \to 0$ (and hence $m \to \infty$), we deduce

$$\lim_{k\to\infty} \int_U \dot\gamma_k(s)\, ds = \int_U \dot\gamma(s)\, ds. \tag{2.18}$$

By approximations, (2.18) holds for all Borel measurable sets $A \subset [a,b]$. and again by further approximations, (2.16) follows. □

2.3.2. Existence of absolutely continuous minimizers. We define a new admissible class as, for fixed $y, z \in \mathbb{R}^n$,

$$\mathcal{A}_{ac} = \{\gamma \in \mathrm{AC}([a,b],\mathbb{R}^n) : \gamma(a) = y, \gamma(b) = z\}.$$

The general framework to obtain existence of minimizers by *the direct method* in calculus of variations goes like the following.

(1) We first show that I is bounded from below, that is, there exists a constant $C > 0$ such that

$$I[\gamma] \ge -C \qquad \text{for all } \gamma \in \mathcal{A}_{ac}.$$

This is usually obtained by the superlinearity of L.

(2) We then take a minimizing sequence $\{\gamma_k\} \subset \mathcal{A}_{ac}$ for I, that is,

$$\lim_{k\to\infty} I[\gamma_k] = \inf_{\mathcal{A}_{ac}} I[\cdot],$$

We use the compactness result (Theorem 2.20) to imply that, upon passing to a subsequence if necessary, $\gamma_k \to \gamma$ uniformly on $[a,b]$ and $\dot\gamma_k \rightharpoonup \dot\gamma$ weakly in $L^1([a,b],\mathbb{R}^n)$.

(3) Finally, we show that $I[\cdot]$ is weakly lower semicontinuous, and in particular,

$$I[\gamma] \le \liminf_{k\to\infty} I[\gamma_k],$$

which yields that $I[\gamma] = \min_{\mathcal{A}_{ac}} I[\cdot]$.

Here is the main result in this section.

Theorem 2.21. *Assume* (2.1). *Then, there exists* $\gamma \in \mathcal{A}_{ac}$ *such that*

$$I[\gamma] = \min_{\mathcal{A}_{ac}} I[\cdot].$$

Proof. We first show that I is bounded from below. From the superlinearity of L in v, for each $\theta > 0$ there exist $C_\theta > 0$ such that

$$L(x,v) \geq \theta|v| - C_\theta \qquad \text{for all } (x,v) \in \mathbb{R}^n \times \mathbb{R}^n. \tag{2.19}$$

In particular, for $\theta = 1$, we have

$$I[\gamma] = \int_a^b L(\gamma(t), \dot\gamma(t))\, dt \geq -(b-a)C_1 \qquad \text{for all } \gamma \in \mathcal{A}_{ac}.$$

Thus, $\inf_{\mathcal{A}_{ac}} I[\cdot]$ is finite, and there exists a sequence $\{\gamma_k\} \subset \mathcal{A}_{ac}$ such that

$$\lim_{k\to\infty} I[\gamma_k] = \inf_{\mathcal{A}_{ac}} I[\cdot],$$

We may assume also that $I[\gamma_k] \leq C$ for all $k \in \mathbb{N}$ for some $C > 0$.

We next show that $\{\gamma_k\}$ satisfies the tightness condition (2.14). Fix $\varepsilon > 0$. Let $E \subset [a,b]$ be a Borel measurable set with measure $|E| < \delta$ with $\delta > 0$ to be chosen. Then,

$$\begin{aligned}\int_E L(\gamma_k(s), \dot\gamma_k(s))\, ds &= I[\gamma_k] - \int_{[a,b]\setminus E} L(\gamma_k(s), \dot\gamma_k(s))\, ds \\ &\leq C + |[a,b] \setminus E|\, C_1 \leq C + (b-a)C_1 \leq C.\end{aligned}$$

We use (2.19) in the above inequality to yield

$$\int_E |\dot\gamma_k(s)| ds \leq \frac{C}{\theta} + \frac{C_\theta |E|}{\theta} \leq \frac{C}{\theta} + \frac{C_\theta \delta}{\theta} \qquad \text{for all } k \in \mathbb{N}.$$

Fix $\theta > 1$ sufficiently large so that $C/\theta < \varepsilon/2$. Then, choose $\delta > 0$ sufficiently small so that $C_\theta \delta/\theta < \varepsilon/2$ to imply (2.14).

We apply Theorem 2.20 to get the existence of $\gamma \in \mathcal{A}_{ac}$ such that, up to passing to a subsequence if needed, $\gamma_k \to \gamma$ uniformly on $[a,b]$, and $\dot\gamma_k \rightharpoonup \dot\gamma$ weakly in $L^1([a,b], \mathbb{R}^n)$. We also have that, for some $C > 0$ and all $k \in \mathbb{N}$,

$$\|\gamma_k\|_{L^\infty([a,b])} + \|\gamma\|_{L^\infty([a,b])} \leq C.$$

We now show that

$$I[\gamma] \leq \liminf_{k\to\infty} I[\gamma_k]. \tag{2.20}$$

To do this, we need the following lemma, whose proof is given later.

Lemma 2.22. *Fix $C, K > 0$ and $\varepsilon > 0$. There exists $\delta > 0$ such that, if $v \in \overline{B}(0,K)$, and $x, y \in \overline{B}(0,C)$ with $|x-y| < \delta$, then*

$$L(y,w) \geq L(x,v) + D_v L(x,v) \cdot (w - v) - \varepsilon \qquad \textit{for all } w \in \mathbb{R}^n.$$

We use the above lemma to prove (2.20). For each $\lambda > 0$, denote by

$$U_\lambda = \{t \in [a,b] \,:\, |\dot\gamma(t)| \leq \lambda\}.$$

As $\lambda \to \infty$, $U_\lambda \to [a,b]$ up to a set of zero Lebesgue measure. In U_λ, for each $\varepsilon > 0$, there exists $\delta > 0$ so that, for $|v| \le \lambda$, $x, y \in \overline{B}(0,C)$ with $|x-y| < \delta$,

$$L(y,w) \ge L(x,v) + D_vL(x,v) \cdot (w-v) - \varepsilon \qquad \text{for all } w \in \mathbb{R}^n.$$

Pick $k \in \mathbb{N}$ large enough to have $\|\gamma_k - \gamma\|_{L^\infty([a,b])} < \delta$. Then, for $s \in U_\lambda$,

$$L(\gamma_k(s), \dot\gamma_k(s)) \ge L(\gamma(s), \dot\gamma(s)) + D_vL(\gamma(s), \dot\gamma(s)) \cdot (\dot\gamma_k - \dot\gamma)(s) - \varepsilon.$$

We compute that, for $k \gg 1$,

$$\begin{aligned} I[\gamma_k] &= \int_{U_\lambda} L(\gamma_k(s), \dot\gamma_k(s))\, ds + \int_{[a,b]\setminus U_\lambda} L(\gamma_k(s), \dot\gamma_k(s))\, ds \\ &\ge \int_{U_\lambda} L(\gamma(s), \dot\gamma(s))\, ds + \int_a^b \mathbf{1}_{U_\lambda}(s) D_vL(\gamma(s), \dot\gamma(s)) \cdot (\dot\gamma_k - \dot\gamma)(s)\, ds \\ &\qquad\qquad - \varepsilon|U_\lambda| - C_1 \left|[a,b] \setminus U_\lambda\right|. \end{aligned}$$

Let $k \to \infty$, $\lambda \to \infty$, and $\varepsilon \to 0$ in this order to get (2.20). We conclude that

$$I[\gamma] = \min_{\mathcal{A}_{ac}} I[\cdot].$$

□

Let us now prove Lemma 2.22.

Proof of Lemma 2.22. By the assumptions, there exists $C > 0$ such that

$$\|L\|_{L^\infty(B(0,C)\times B(0,K))} + \|D_vL\|_{L^\infty(B(0,C)\times B(0,K))} \le C.$$

Thus, for $|x| \le C$ and $|v| \le K$,

$$L(x,v) + D_vL(x,v) \cdot (w-v) - \varepsilon \le C + C|w-v| \le C(1+|w|).$$

As $L(y,w)$ is superlinear in w, there exists $C_1 > 0$ such that, for $|w| \ge C_1$,

$$L(y,w) \ge C(2+|w|) \ge L(x,v) + D_vL(x,v) \cdot (w-v) - \varepsilon.$$

Hence, the needed inequality holds automatically if $|w| \ge C_1$. We then only need to consider the case $|w| \le C_1$. As L is uniformly continuous on $\overline{B}(0,C) \times \overline{B}(0,C_1)$, there exists $\delta > 0$ such that, for $|w| \le C_1$ and $x, y \in \overline{B}(0,C)$ with $|x-y| < \delta$,

$$|L(y,w) - L(x,w)| \le \varepsilon.$$

Thus, for $|w| \le C_1$ and $x, y \in \overline{B}(0,C)$ with $|x-y| < \delta$,

$$L(y,w) \ge L(x,w) - \varepsilon \ge L(x,w) + D_vL(x,v) \cdot (w-v) - \varepsilon.$$

The proof is complete.

□

2.3.3. Regularity of absolutely continuous minimizers. It turns out that absolutely continuous minimizers are also C^k.

Theorem 2.23. *Assume* (2.1). *Let $\gamma \in \mathcal{A}_{ac}$ be a minimizer of the action for L in the admissible class $\mathcal{A}_{ac}$. Then, $\gamma \in C^k$.*

The proof of this theorem is rather complicated and involved. An important ingredient in the proof that we use is the method of characteristics to give local existence of solutions to Hamilton–Jacobi equations. We need the following important lemma.

Lemma 2.24. *Let $\pi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ be the projection map with $\pi(x,v) = v$ for all $(x,v) \in \mathbb{R}^n \times \mathbb{R}^n$. Let ϕ_t^L be the Lagrangian flow. Fix $C > 0$. Then, there exists $\delta_1 > 0$ such that*

$$\pi \circ \phi_s^L(\{x\} \times B(0,2C)) \supset B(x, C|s|) \quad \textit{for all } |s| \le \delta_1. \tag{2.21}$$

Proof. For (x,v), recall that $\phi_t^L(x,v)$ is the Lagrangian flow at time t. We write

$$\phi_t^L(x,v) = \left(\gamma(t,x,v), \frac{\partial \gamma}{\partial t}(t,x,v)\right) = (\gamma(t), \dot{\gamma}(t)).$$

Of course, $(\gamma(0,x,v), \frac{\partial \gamma}{\partial t}(0,x,v)) = (x,v)$. Let

$$\Gamma(t,x,v) = \frac{\gamma(t,x,v) - \gamma(0,x,v)}{t} = \int_0^1 \frac{\partial \gamma}{\partial t}(st,x,v)\, ds.$$

Then,

$$\begin{aligned} \frac{\partial \Gamma}{\partial t}(t,x,v) &= \int_0^1 \frac{\partial^2 \gamma}{\partial t^2}(st,x,v) s\, ds, \\ \frac{\partial \Gamma}{\partial v}(t,x,v) &= \int_0^1 \frac{\partial^2 \gamma}{\partial v \partial t}(st,x,v)\, ds. \end{aligned}$$

Let $\tilde{\Gamma} : \mathbb{R} \times B(0,2C) \to \mathbb{R} \times \mathbb{R}^n$ be such that

$$\tilde{\Gamma}(t,v) = (t, \Gamma(t,x,v)).$$

We see that

$$D_{(t,v)}\tilde{\Gamma}(0,v) = \begin{pmatrix} 1 & 0 \\ \frac{\partial \Gamma}{\partial t}(0,x,v) & I_n \end{pmatrix}.$$

As $\det D_{(t,v)}\tilde{\Gamma}(0,v) = 1$, we use the inverse mapping theorem to deduce that, for each $v \in B(0, 3C/2)$, there exist $\delta_v > 0$, $r_v, l_v \in (0, C/2)$, and an open set O_v such that

$$\tilde{\Gamma} : (-\delta_v, \delta_v) \times B(v, r_v) \to O_v$$

is a C^1 diffeomorphism, and $(-\delta_v, \delta_v) \times B(v, l_v) \subset O_v$.

As $\{B(v,l_v)\}_{v\in B(0,3C/2)}$ is an open cover of $B(0,C)$, we can find $v_1,\ldots,v_k\in B(0,3C/2)$ such that

$$B(0,C)\subset\bigcup_{i=1}^k B(v_i,l_{v_i}).$$

Let $\delta_1=\min_{1\le i\le k}\delta_{v_i}$. Then,

$$\tilde\Gamma\left((-\delta_1,\delta_1)\times\bigcup_{i=1}^k B(v_i,r_{v_i})\right)\supset(-\delta_1,\delta_1)\times B(0,C),$$

which implies (2.21).

□

Proof of Theorem 2.23. We divide the proof into several steps.

Step 1. As L is superlinear, we have $L(x,v)\ge|v|-C_1$ for all $(x,v)\in\mathbb{R}^n\times\mathbb{R}^n$ (see (2.19)). In particular,

$$\int_a^b|\dot\gamma(s)|\,ds\le\int_a^b L(\gamma(s),\dot\gamma(s))\,ds+C_1(b-a)\le C.$$

We use the above and the fact that $\dot\gamma$ exists almost everywhere in (a,b) to yield the existence of $t_0\in(a,b)$ such that $\dot\gamma(t_0)$ exists and $|\dot\gamma(t_0)|\le C/2$ for some $C>0$. By using the definition of differentiability of γ at t_0, there exists $\delta>0$ sufficiently small such that

$$|\gamma(t)-\gamma(t_0)|\le C|t-t_0|\qquad\text{for } t\in(t_0-\delta,t_0+\delta). \tag{2.22}$$

Without loss of generality, we assume $t_0=0$.

Step 2. Write $\gamma(0)=x_0$. By (2.21), there exists $\delta_1>0$ such that

$$\pi\circ\phi_s^L(\{x_0\}\times B(0,2C))\supset B(x_0,C|s|)\qquad\text{for all } |s|\le\delta_1. \tag{2.23}$$

We can choose $\delta_1<\delta$.

Step 3. For each $v\in B(0,2C)$, let $p=D_vL(x_0,v)$. Choose $\psi_v\in C^2(\mathbb{R}^n)$ such that $\|\psi_v\|_{C^2(\mathbb{R}^n)}\le 4(C+1)$, $D\psi_v(x_0)=p$, and $\operatorname{supp}(\psi_v)\subset B(0,R)$ for some $R>0$ sufficiently large independent of v.

By the method of characteristics and finite speed of propagations, we have a local wellposedness for the following Hamilton–Jacobi equation

$$\begin{cases}u_t^v+H(x,Du^v)=0 & \text{in } \mathbb{R}^n\times(-\delta_2,\delta_2),\\ u^v(x,0)=\psi_v(x) & \text{on } \mathbb{R}^n,\end{cases} \tag{2.24}$$

where $\delta_2\in(0,\delta_1)$ is a small positive number independent of v, and

$$u^v\in C^2(B(0,2R)\times(-\delta_2,\delta_2)).$$

It is important to note that we only need to have local solvability of the solution u^v via the method of characteristics.

Let $(X(t), P(t))$ be the corresponding Hamiltonian system with

$$(X(0), P(0)) = (x_0, D\psi_v(x_0)) = (x_0, p).$$

Step 4. In light of (2.22) and (2.23), there exists $v \in B(0, 2C)$ such that

$$\gamma(\delta_2) = \pi \circ \phi^L_{\delta_2}(x_0, v).$$

Pick any $\eta \in \mathrm{AC}([0, \delta_2], \mathbb{R}^n)$ with $\eta(0) = \gamma(0) = x_0$ and $\eta(\delta_2) = \gamma(\delta_2)$. Then,

$$\begin{aligned}
&u^v(\eta(\delta_2), \delta_2) - u^v(x_0, 0)\\
&= \int_0^{\delta_2} [Du^v(\eta(s), s) \cdot \dot\eta(s) + u^v_t(\eta(s), s)]\, ds\\
&\le \int_0^{\delta_2} [H(\eta(s), Du^v(\eta(s), s)) + L(\eta(s), \dot\eta(s)) + u^v_t(\eta(s), s)]\, ds\\
&\le \int_0^{\delta_2} L(\eta(s), \dot\eta(s))\, ds.
\end{aligned}$$

Thus,

$$u^v(\eta(\delta_2), \delta_2) - u^v(x_0, 0) \le \int_0^{\delta_2} L(\eta(s), \dot\eta(s))\, ds. \tag{2.25}$$

Step 5. On the other hand, write

$$\alpha(s) = \pi \circ \phi^L_s(x_0, v) \quad \text{for } s \in [0, \delta_2].$$

Then, for $s \in [0, \delta_2]$,

$$\dot\alpha(s) = \dot X(s) = D_pH(\alpha(s), Du^v(\alpha(s), s)). \tag{2.26}$$

From (2.25) and (2.26), we see that the inequality in (2.25) becomes an equality if and only if $\eta = \alpha$. Therefore, α is the unique minimizer of the action of L on $[0, \delta_2]$, which means that $\gamma = \alpha$ on $[0, \delta_2]$.

Thus, γ is C^k in $(0, \delta_2)$. It is then easily seen that $\gamma \in C^k$ and γ solves the Euler-Lagrange equations

$$\begin{cases} \frac{d}{dt}\left(D_vL(\gamma(t), \dot\gamma(t))\right) = D_xL(\gamma(t), \dot\gamma(t)) & \text{for } t \in [a, b],\\ \gamma(0) = x, \dot\gamma(0) = v. \end{cases}$$

$\square$

2.4. References

(1) The content in this chapter is classical. We refer the readers to Cannarsa and Sinestrari [**CS04**], and Fathi [**Fat**] for similar materials.

(2) For the method of characteristics for Hamilton–Jacobi equations, see Appendix C. See also Evans [**Eva10**, Chapter 3].

Chapter 3

Hamilton–Jacobi equations on a torus

In this chapter, we always consider a given Hamiltonian $H : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ that satisfies

$$(3.1) \qquad \begin{cases} H \in C(\mathbb{R}^n \times \mathbb{R}^n), \\ y \mapsto H(y,p) \text{ is } \mathbb{Z}^n\text{-periodic}, \\ \lim_{|p|\to\infty} \min_{y\in\mathbb{R}^n} H(y,p) = +\infty. \end{cases}$$

Here, $y \mapsto H(y,p)$ is $\mathbb{Z}^n$-periodic means that, for $(y,p) \in \mathbb{R}^n \times \mathbb{R}^n$ and $k \in \mathbb{Z}^n$,

$$H(y,p) = H(y+k,p).$$

Let $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ be the usual flat n-dimensional torus. Then, we can think of $H \in C(\mathbb{T}^n \times \mathbb{R}^n)$. The third condition in (3.1) is often called the coercivity condition.

3.1. Cell problems

For each fixed $p \in \mathbb{R}^n$, the cell problem of interests is

$$(3.2) \qquad H(y, p + Dv(y)) = c \qquad \text{in } \mathbb{T}^n.$$

Here, we search for a pair of unknowns $(v,c) \in C(\mathbb{T}^n) \times \mathbb{R}$, where v solves (3.2) in the viscosity sense. The equation (3.2) is also called ergodic problem or corrector problem in the literature, and it plays the essential role in many different fields. In fact, it is the main object of this book as we will see later.

The main theorem of this section is the following.

Theorem 3.1. *Assume* (3.1). *Fix* $p \in \mathbb{R}^n$. *Then, there exists a unique constant* $c \in \mathbb{R}$ *such that the cell problem* (3.2) *has a viscosity solution* $v \in C(\mathbb{T}^n)$.

Proof. For each $\lambda > 0$, consider the following approximated equation

$$\lambda v^\lambda + H(y, p + Dv^\lambda) = 0 \qquad \text{in } \mathbb{R}^n. \tag{3.3}$$

Set $C_0 = \max_{y \in \mathbb{T}^n} |H(y,p)|$. It is clear that $-C_0/\lambda$ is a subsolution to (3.3), and C_0/λ is a supersolution to (3.3). Then, by the wellposedness of viscosity solutions to static Hamilton–Jacobi equations (see [**Tra21**, Chapter 1]), (3.3) has a unique solution $v^\lambda \in C(\mathbb{R}^n)$, and

$$-\frac{C_0}{\lambda} \leq v^\lambda \leq \frac{C_0}{\lambda}. \tag{3.4}$$

By the $\mathbb{Z}^n$-periodicity of H in y, we see that $v^\lambda(\cdot + k)$ is also a solution to (3.3) for each $k \in \mathbb{Z}^n$. Then, the uniqueness of solutions to (3.3) yields that $v^\lambda = v^\lambda(\cdot + k)$ for each $k \in \mathbb{Z}^n$. Thus, v^λ is $\mathbb{Z}^n$-periodic, or we write $v^\lambda \in C(\mathbb{T}^n)$. By the bound (3.4) and the coercivity of H,

$$H(y, p + Dv^\lambda) \leq C_0 \quad \Longrightarrow \quad \|Dv^\lambda\|_{L^\infty(\mathbb{T}^n)} \leq C. \tag{3.5}$$

Combining (3.4) and (3.5), we deduce that there exists $C > 0$ independent of $\lambda > 0$ such that

$$\lambda \|v^\lambda\|_{L^\infty(\mathbb{T}^n)} + \|Dv^\lambda\|_{L^\infty(\mathbb{T}^n)} \leq C. \tag{3.6}$$

Let $w^\lambda(y) = v^\lambda(y) - v^\lambda(0)$ for $y \in \mathbb{T}^n$. Then, thanks to (3.6),

$$\|w^\lambda\|_{L^\infty(\mathbb{T}^n)} + \|Dw^\lambda\|_{L^\infty(\mathbb{T}^n)} \leq \sqrt{n}\|Dv^\lambda\|_{L^\infty(\mathbb{T}^n)} + \|Dv^\lambda\|_{L^\infty(\mathbb{T}^n)} \leq C.$$

By Arzelà-Ascoli's theorem, there exists a sequence $\{\lambda_j\} \to 0$ such that

$$\begin{cases} w^{\lambda_j} \to v & \text{uniformly on } \mathbb{T}^n, \\ \lambda_j v^{\lambda_j} \to -c & \text{uniformly on } \mathbb{T}^n, \end{cases}$$

for some $v \in \mathrm{Lip}\,(\mathbb{T}^n)$ and $c \in \mathbb{R}$. Besides, in light of (3.3), w^{λ_j} solves

$$\lambda_j w^{\lambda_j} + H(y, p + Dw^{\lambda_j}) = -\lambda_j v^{\lambda_j}(0) \qquad \text{in } \mathbb{T}^n.$$

Let $j \to \infty$ and use the stability result for viscosity solutions to deduce

$$H(y, p + Dv) = c \qquad \text{in } \mathbb{T}^n.$$

We have thus obtained the existence of a pair $(v, c) \in C(\mathbb{T}^n) \times \mathbb{R}$ solving the above, which is exactly (3.2).

It remains to show that c is unique. Assume otherwise that there exist two pairs $(v_1, c_1), (v_2, c_2) \in C(\mathbb{T}^n) \times \mathbb{R}$ solving (3.2) with $c_1 < c_2$. We pick $\delta > 0$ sufficiently small so that, in the viscosity sense,

$$\delta v_1 + H(y, p + Dv_1) \leq \frac{c_1 + c_2}{2} \leq \delta v_2 + H(y, p + Dv_2) \qquad \text{in } \mathbb{T}^n.$$

By the comparison principle for the above equation, we yield that $v_1 \le v_2$. The same logic gives $v_1 + C \le v_2$ for any constant $C > 0$, which is absurd. Hence, we must have that c is unique. □

Definition 3.2 (Effective Hamiltonian)**.** Assume (3.1). For $p \in \mathbb{R}^n$, let $c \in \mathbb{R}$ be the unique constant such that the cell problem (3.2) has a viscosity solution $v \in C(\mathbb{T}^n)$. Denote by $\overline{H}(p) = c$.

We say that $\overline{H}$ is the effective Hamiltonian corresponding to H.

The cell problem (3.2) now can be written as

$$H(y, p + Dv(y)) = \overline{H}(p) \qquad \text{in } \mathbb{T}^n. \tag{3.7}$$

In the literature, we also say that $\overline{H}(p)$ is the additive eigenvalue of (3.7). It is rather clear to see that $\overline{H}(p)$ is defined in a very implicit way, and it is not easy at all to read off information of $\overline{H}$. In fact, understanding fine properties of the effective Hamiltonian is a central goal in both PDE and weak KAM theory.

Remark 3.3. Although $\overline{H}(p)$ is uniquely defined, viscosity solutions to (3.7) are not unique. Clearly, if v is a solution, then $v + C$ is also a solution for any $C \in \mathbb{R}$. We will see later that there could be many more other solutions, and we will characterize solutions in terms of uniqueness sets.

Whenever needed, we write $v(y) = v(y, p)$ or $v(y) = v_p(y)$ to demonstrate clearly the dependence of v on p.

The following lemma is quite straightforward.

Lemma 3.4. *Assume* (3.1)*. Then,*

$$\min_{y \in \mathbb{T}^n} H(y, p) \le \overline{H}(p) \le \max_{y \in \mathbb{T}^n} H(y, p).$$

In particular, $\overline{H}$ is coercive.

Proof. We only need to prove the first inequality as the second one follows in a similar manner. Pick $x_1 \in \mathbb{T}^n$ such that $v(x_1) = \max_{\mathbb{T}^n} v$. Then, the constant function $\phi \equiv v(x_1)$ touches v from above at x_1, and by the viscosity subsolution test,

$$H(x_1, p + D\phi(x_1)) = H(x_1, p) \le \overline{H}(p) \quad \Longrightarrow \quad \min_{y \in \mathbb{T}^n} H(y, p) \le \overline{H}(p).$$

□

We next show that $\overline{H}$ is continuous.

Lemma 3.5. *Assume* (3.1)*. Then, $\overline{H}$ is continuous.*

Proof. Let $\{p_k\}$ be a sequence in $\mathbb{R}^n$ convergence to p. There exists $C > 0$ such that $|p_k| \le C$ for all $k \in \mathbb{N}$. Let v_k be a solution to the cell problem with respect to p_k such that $v_k(0) = 0$ after adding a constant if needed. Of course, v_k solves

$$H(y, p_k + Dv_k(y)) = \overline{H}(p_k) \qquad \text{in } \mathbb{T}^n. \tag{3.8}$$

By the coercivity of H, and the points that $|p_k| \le C$, $v_k(0) = 0$, we get that

$$\|v_k\|_{L^\infty(\mathbb{T}^n)} + \|Dv_k\|_{L^\infty(\mathbb{T}^n)} \le C.$$

By the Arzelà-Ascoli theorem and by passing to a subsequence if needed, we may assume that $v_k \to v$ uniformly on $\mathbb{T}^n$ for some $v \in \mathrm{Lip}\,(\mathbb{T}^n)$, and $\overline{H}(p_k) \to c$ for some $c \in \mathbb{R}$. By the usual stability result for viscosity solutions, we see that v solves

$$H(y, p + Dv(y)) = c \qquad \text{in } \mathbb{T}^n.$$

Thus, $\overline{H}(p) = c$, and we conclude that $\overline{H}$ is continuous. □

Qualitatively, we have that $\overline{H}$ is continuous and coercive. It is however very hard to study finer properties of $\overline{H}$ in this very general setting. For example, it is not clear at all what are the relations between the level sets of $\overline{H}$ and H.

3.2. Large time averages

In this section, we show that the effective Hamiltonian can be obtained via large time averages of solutions to Cauchy problem.

Theorem 3.6. *Assume* (3.1). *Fix $p \in \mathbb{R}^n$. Let u be the viscosity solution to*

$$\begin{cases} u_t + H(y, p + Du) = 0 & \text{in } \mathbb{T}^n \times (0,\infty), \\ u(y,0) = 0 & \text{on } \mathbb{T}^n. \end{cases} \tag{3.9}$$

Then,

$$\lim_{t\to\infty} \frac{u(y,t)}{t} = -\overline{H}(p).$$

In fact,

$$\left| \frac{u(y,t)}{t} + \overline{H}(p) \right| \le \frac{C}{t} \qquad \text{for all } (y,t) \in \mathbb{T}^n \times (0,\infty),$$

where $C = C(p) > 0$ is a constant.

Proof. Let v be a solution to the cell problem (3.7) such that $v(0) = 0$. Then, by coercivity of H,

$$\|v\|_{L^\infty(\mathbb{T}^n)} \le \sqrt{n}\|Dv\|_{L^\infty(\mathbb{T}^n)} \le C,$$

for some $C = C(p) > 0$. Denote by

$$\varphi^{\pm}(y,t) = v(y) \pm \|v\|_{L^\infty(\mathbb{T}^n)} - \overline{H}(p)t \qquad \text{for all } (y,t) \in \mathbb{T}^n \times [0,\infty).$$

It is clear that $\varphi^-(y,0) \le 0 \le \varphi^+(y,0)$ for $y \in \mathbb{T}^n$, and $\varphi^\pm$ are both solutions to

$$u_t + H(y, p + Du) = 0 \qquad \text{in } \mathbb{T}^n \times (0,\infty).$$

Thus, φ^- and φ^+ are a viscosity subsolution and a viscosity supersolution to (3.9), respectively. By the comparison principle for (3.9),

$$\varphi^- \le u \le \varphi^+ \quad \Longrightarrow \quad \left| \frac{u(y,t)}{t} + \overline{H}(p) \right| \le \frac{2\|v\|_{L^\infty(\mathbb{T}^n)}}{t} \le \frac{C}{t}.$$

□

Remark 3.7. In various situations, we actually have explicitly the dependence of $C = C(p)$ on p in the above theorem once we know the growth condition of H. For example, assume that there exists $C_0 > 0$ such that

$$\frac{1}{2}|p|^2 - C_0 \le H(y,p) \le \frac{1}{2}|p|^2 + C_0 \qquad \text{for all } (y,p) \in \mathbb{T}^n \times \mathbb{R}^n.$$

Then, by Lemma 3.4, $\overline{H}$ also enjoys the above bounds. Let v be a solution to the cell problem (3.7) such that $v(0) = 0$. We have

$$\frac{1}{2}|p + Dv|^2 - C_0 \le \frac{1}{2}|p|^2 + C_0 \quad \Longrightarrow \quad |p + Dv| \le |p| + 4C_0.$$

Thus,

$$|Dv| \le 2|p| + 4C_0 \quad \Longrightarrow \quad C(p) = 4\sqrt{n}(|p| + 2C_0).$$

3.3. Effective Hamiltonians in the convex setting

We now impose more conditions on H. Throughout this section, we assume that

$$p \mapsto H(y,p) \quad \text{is convex for each } y \in \mathbb{T}^n. \tag{3.10}$$

Theorem 3.8. *Assume* (3.1) *and* (3.12). *Then, for each $p \in \mathbb{R}^n$,*

$$\overline{H}(p) = \inf_{\phi \in C^1(\mathbb{T}^n)} \max_{y \in \mathbb{T}^n} H(y, p + D\phi(y)).$$

Proof. Fix $p \in \mathbb{R}^n$. For each $\phi \in C^1(\mathbb{T}^n)$, denote by

$$c_\phi = \max_{y \in \mathbb{T}^n} H(y, p + D\phi(y)).$$

By repeating the proof of the uniqueness of $\overline{H}(p)$ (the last part of the proof of Theorem 3.1), we deduce that $c_\phi \ge \overline{H}(p)$. Hence,

$$\inf_{\phi \in C^1(\mathbb{T}^n)} \max_{y \in \mathbb{T}^n} H(y, p + D\phi(y)) = \inf_{\phi \in C^1(\mathbb{T}^n)} c_\phi \ge \overline{H}(p).$$

We now prove the reverse inequality. Let v be a solution to the cell problem (3.7). Then, $v \in \mathrm{Lip}\,(\mathbb{T}^n)$, and v satisfies

$$H(y, p + Dv(y)) \leq \overline{H}(p) \qquad \text{for a.e. } y \in \mathbb{T}^n. \tag{3.11}$$

We now use convolutions with standard mollifiers to smooth v up. Take η to be a standard mollifier, that is,

$$\eta \in C_c^\infty(\mathbb{R}^n, [0,\infty)), \qquad \mathrm{supp}(\eta) \subset B(0,1), \qquad \int_{\mathbb{R}^n} \eta(x)\,dx = 1.$$

For $\varepsilon > 0$, denote by $\eta_\varepsilon(x) = \varepsilon^{-n}\eta\left(\frac{x}{\varepsilon}\right)$ for all $x \in \mathbb{R}^n$. For $x \in \mathbb{R}^n$, denote by

$$v^\varepsilon(x) = (\eta_\varepsilon \star v)(x) = \int_{\mathbb{R}^n} \eta_\varepsilon(x-y)v(y)\,dy = \int_{B(x,\varepsilon)} \eta_\varepsilon(x-y)v(y)\,dy.$$

Then $v^\varepsilon \in C^\infty(\mathbb{R}^n)$, v^ε is $\mathbb{Z}^n$-periodic, and $v^\varepsilon \to v$ locally uniformly as $\varepsilon \to 0$. Let

$$\omega(\varepsilon) = \max\left\{|H(x,q) - H(y,q)| \,:\, x \in \mathbb{T}^n, y \in B(x,\varepsilon), H(x,q) \leq \overline{H}(p)\right\}.$$

Of course, $\lim_{\varepsilon\to 0}\omega(\varepsilon) = 0$. We use (3.11), the above, and Jensen's inequality to yield that

$$\begin{aligned}
\overline{H}(p) &\geq \int_{B(x,\varepsilon)} \eta_\varepsilon(x-y)H(y, p + Dv(y))\,dy \\
&\geq \int_{B(x,\varepsilon)} \eta_\varepsilon(x-y)H(x, p + Dv(y))\,dy - \omega(\varepsilon) \\
&\geq H\left(x, p + \int_{B(x,\varepsilon)} \eta_\varepsilon(x-y)Dv(y)\,dy\right) - \omega(\varepsilon) \\
&\geq H(x, Dv^\varepsilon(x)) - \omega(\varepsilon).
\end{aligned}$$

Thus, v^ε satisfies

$$H(x, Dv^\varepsilon(x)) \leq \overline{H}(p) + \omega(\varepsilon) \qquad \text{for all } x \in \mathbb{T}^n.$$

Let $\varepsilon \to 0$ to conclude. □

By using the inf-max formula, we have rather immediately the following result.

Theorem 3.9. *Assume* (3.1) *and* (3.12). *Then,* $\overline{H}$ *is convex.*

Proof. Fix $p_1, p_2 \in \mathbb{R}^n$. For each $\phi \in C^1(\mathbb{T}^n)$,

$$\begin{aligned}
&\max_{y\in\mathbb{T}^n} H\left(y, \frac{p_1+p_2}{2} + D\phi(y)\right) \\
&\le \frac{1}{2}\left(\max_{y\in\mathbb{T}^n} H(y, p_1 + D\phi(y)) + \max_{y\in\mathbb{T}^n} H(y, p_2 + D\phi(y))\right) \\
&\le \frac{1}{2}\left(\overline{H}(p_1) + \overline{H}(p_2)\right).
\end{aligned}$$

Take infimum over $\phi \in C^1(\mathbb{T}^n)$ to conclude. □

3.4. Backward characteristics in the convex setting

In this section, we assume

$$\begin{cases} H \in C^k(\mathbb{T}^n \times \mathbb{R}^n) \text{ for some } k \ge 2, \\ D^2_{pp}H(y,p) > 0 \text{ for all } (y,p) \in \mathbb{T}^n \times \mathbb{R}^n, \\ \lim_{|p|\to\infty} \min_{y\in\mathbb{T}^n} \frac{H(y,p)}{|p|} = +\infty. \end{cases} \tag{3.12}$$

Let v be a solution to the cell problem (3.7), that is, v solves

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

Then, $u(y,t) = p \cdot y + v(y) - \overline{H}(p)t$ is the unique viscosity solution to

$$\begin{cases} u_t + H(y, Du) = 0 & \text{in } \mathbb{R}^n \times (0,\infty), \\ u(y,0) = p \cdot y + v(y) & \text{on } \mathbb{R}^n. \end{cases} \tag{3.13}$$

We now use the optimal control formula for u to get the backward characteristics for v.

Theorem 3.10. *Assume* (3.12)*. For $p \in \mathbb{R}^n$, let v be a solution to the cell problem* (3.7)*. Then, for each $x \in \mathbb{R}^n$, there exists a C^k curve $\xi : (-\infty, 0] \to \mathbb{R}^n$ such that $\xi(0) = x$, and for all $t_2 \le t_1 \le 0$,*

$$p \cdot \xi(t_1) + v(\xi(t_1)) - p \cdot \xi(t_2) - v(\xi(t_2)) = \int_{t_2}^{t_1} \left(L(\xi(s), \dot{\xi}(s)) + \overline{H}(p)\right) ds.$$

Moreover,

$$\|\dot{\xi}\|_{L^\infty((-\infty,0])} \le C$$

for some $C = C(|p|) > 0$.

Proof. As noted above, $u(y,t) = p \cdot y + v(y) - \overline{H}(p)t$ is the unique viscosity solution to (3.13). We construct ξ on $[-m, -m+1]$ iteratively for $m \in \mathbb{N}$ as follows.

We are given that $\xi(0) = x$. For $m \in \mathbb{N}$, by the optimal control formula,

$$u(\xi(-m+1),1) = \inf\Big\{ \int_0^1 L(\gamma(s),\dot\gamma(s))\,ds + u(\gamma(0),0)\ : \\ \gamma \in \mathrm{AC}\,([0,1],\mathbb{R}^n), \gamma(1) = \xi(-m+1)\Big\}.$$

Write

$$I[\gamma] = \int_0^1 L(\gamma(s),\dot\gamma(s))\,ds.$$

Pick $\theta > |p|+1$. As L is superlinear in v, there exists $C_\theta > 0$ such that

$$L(y,v) \geq \theta|v| - C_\theta \qquad \text{for all } (y,v) \in \mathbb{R}^n \times \mathbb{R}^n.$$

Hence, for each $\gamma \in \mathrm{AC}\,([0,1],\mathbb{R}^n)$ with $\gamma(1) = \xi(-m+1)$ fixed,

$$\begin{aligned} I[\gamma] + u(\gamma(0),0) &= \int_0^1 L(\gamma(s),\dot\gamma(s))\,ds + u(\gamma(0),0) \\ &\geq \int_0^1 (\theta|\dot\gamma(s)| - C_\theta)\ ds + p\cdot\gamma(0) + v(\gamma(0)) \\ &\geq \theta|\gamma(1)-\gamma(0)| + p\cdot\gamma(0) + v(\gamma(0)) - C_\theta \\ &\geq (\theta - |p|)|\gamma(0)| - C \geq |\gamma(0)| - C. \end{aligned}$$

Thus, we can find $R > 0$ sufficiently large such that

$$\begin{aligned} &\inf\{I[\gamma] + u(\gamma(0),0)\ :\ \gamma \in \mathrm{AC}\,([0,1]), \gamma(1) = \xi(-m+1)\} \\ &= \inf\{I[\gamma] + u(\gamma(0),0)\ :\ \gamma \in \mathrm{AC}\,([0,1]), \gamma(1) = \xi(-m+1), |\gamma(0)| \leq R\}. \end{aligned}$$

For each $y \in \overline{B}(0,R)$, let

$$Q(y) = \inf\{I[\gamma]\ :\ \gamma \in \mathrm{AC}\,([0,1]), \gamma(1) = \xi(-m+1), \gamma(0) = y\}.$$

By Theorems 2.21 and 2.23, there exists $\gamma^y \in C^k([0,1])$ with $\gamma^y(1) = \xi(-m+1), \gamma^y(0) = y$ such that

$$Q(y) = I[\gamma^y].$$

Moreover, the proof of Theorem 2.21 also gives us that Q is lower semicontinuous. Therefore, there exist $z \in \overline{B}(0,R)$ and $\eta \in C^k([0,1])$ with $\eta(1) = \xi(-m+1)$, $\eta(0) = z$ such that

$$\begin{aligned} &u(\xi(-m+1),1) \\ &= \inf\{I[\gamma] + u(\gamma(0),0)\ :\ \gamma \in \mathrm{AC}\,([0,1]), \gamma(1) = \xi(-m+1), |\gamma(0)| \leq R\} \\ &= \inf\{Q(y) + p\cdot y + v(y)\ :\ |y| \leq R\} \\ &= Q(z) + p\cdot z + v(z) = I[\eta] + p\cdot z + v(z). \end{aligned}$$

Denote

$$\xi(-k+s) = \eta(s) \qquad \text{for } s \in [0,1].$$

By this iterative way, we obtained that ξ is defined on $(-\infty, 0]$ with $\xi(0) = x$. Furthermore, by the Dynamic Programming Principle, for $t \in (0,1)$,

$$\begin{aligned} u(\xi(-m+1),1) = \inf \Big\{ \int_t^1 & L(\gamma(s),\dot\gamma(s))\,ds + u(\gamma(t),t) \ : \\ & \gamma \in \mathrm{AC}\,([t,1],\mathbb{R}^n), \gamma(1) = \xi(-m+1) \Big\} \\ = \int_{-m+t}^{-m+1} & L(\xi(s),\dot\xi(s))\,ds + u(\xi(-m+t),t). \end{aligned}$$

Thus, by using the definition of u that $u(y,t) = p \cdot y + v(y) - \overline{H}(p)t$, we imply that, for all $t_2 \le t_1 \le 0$,

$$p \cdot \xi(t_1) + v(\xi(t_1)) - p \cdot \xi(t_2) - v(\xi(t_2)) = \int_{t_2}^{t_1} \Big(L(\xi(s),\dot\xi(s)) + \overline{H}(p) \Big)\,ds,$$

and $\xi \in C^k((-\infty,0])$.

Finally, by the fact that u is differentiable along backward characteristics, we obtain that v is differentiable on $\xi(s)$ for $s \in (-\infty,0)$, and

$$\dot\xi(s) = D_pH(\xi(s), p + Dv(\xi(s))).$$

Therefore,

$$\|\dot\xi\|_{L^\infty((-\infty,0])} \le C$$

for some $C = C(|p|) > 0$. The proof is complete. □

Remark 3.11. By approximations, we see that backward characteristics exist under a weaker condition than (3.12). Indeed, we only need to assume

$$\begin{cases} H \in C(\mathbb{T}^n \times \mathbb{R}^n), \\ p \mapsto H(y,p) \text{ is convex for each } y \in \mathbb{T}^n, \\ \lim_{|p|\to\infty} \min_{y\in\mathbb{T}^n} \frac{H(y,p)}{|p|} = +\infty \end{cases} \tag{3.14}$$

to have existence of Lipschitz backward characteristics. Of course, under (3.14), we would not have the C^k regularity of these curves.

Definition 3.12 (Backward characteristics)**.** Assume either (3.12) or (3.14). For each $x \in \mathbb{R}^n$, let $\xi : (-\infty,0] \to \mathbb{R}^n$ be a Lipschitz curve such that $\xi(0) = x$, and for all $t_2 \le t_1 \le 0$,

$$p \cdot \xi(t_1) + v(\xi(t_1)) - p \cdot \xi(t_2) - v(\xi(t_2)) = \int_{t_2}^{t_1} \Big(L(\xi(s),\dot\xi(s)) + \overline{H}(p) \Big)\,ds.$$

We say that ξ is a backward characteristic of v emanating from x.

Let us also give the definition of global characteristics here.

Definition 3.13 (Global characteristics)**.** Assume either (3.12) or (3.14). If $\xi : \mathbb{R} \to \mathbb{R}^n$ is a Lipschitz curve satisfying that, for all $t_2 \le t_1$,

$$p \cdot \xi(t_1) + v(\xi(t_1)) - p \cdot \xi(t_2) - v(\xi(t_2)) = \int_{t_2}^{t_1} \Big(L(\xi(s), \dot{\xi}(s)) + \overline{H}(p) \Big) \, ds,$$

then we say that ξ is a global characteristic of v.

On the other hand, for arbitrary Lipschitz curves, we always have the following one-sided control.

Lemma 3.14. *Assume* (3.12). *For $p \in \mathbb{R}^n$, let v be a Lipschitz viscosity subsolution to the cell problem* (3.7). *Let $\gamma : (-\infty, 0] \to \mathbb{R}^n$ be an arbitrary Lipschitz curve. Then, for every $T > 0$,*

$$\int_{-T}^{0} \big(L(\gamma(t), \dot{\gamma}(t)) + \overline{H}(p) \big) \, dt \ge p \cdot \gamma(0) + v(\gamma(0)) - p \cdot \gamma(-T) - v(\gamma(-T)).$$

If everything is smooth, then this result is not hard to prove. Here is a quick proof:

$$\begin{aligned} &\int_{-T}^{0} \big(L(\gamma(t), \dot{\gamma}(t)) + \overline{H}(p) \big) \, dt \\ = &\int_{-T}^{0} \left(L(\gamma(t), \dot{\gamma}(t)) + H(\gamma(t), p + Dv(\gamma(t))) \right) dt \\ \ge &\int_{-T}^{0} \dot{\gamma}(t) \cdot (p + Dv(\gamma(t))) \, dt = p \cdot \gamma(0) + v(\gamma(0)) - p \cdot \gamma(-T) - v(\gamma(-T)). \end{aligned}$$

As v is only Lipschitz in general, the above computation is only heuristic. To overcome this difficulty, we perform a convolution trick to smooth v up.

Proof of Lemma 3.14. Take η to be the standard mollifier, that is,

$$\eta \in C_c^\infty(\mathbb{R}^n, [0, \infty)), \qquad \operatorname{supp}(\eta) \subset B(0,1), \qquad \int_{\mathbb{R}^n} \eta(x) \, dx = 1.$$

For $\varepsilon > 0$, denote by $\eta_\varepsilon(x) = \varepsilon^{-n} \eta\left(\frac{x}{\varepsilon}\right)$ for all $x \in \mathbb{R}^n$. Set, for $x \in \mathbb{R}^n$,

$$v^\varepsilon(x) = (\eta_\varepsilon \star v)(x) = \int_{\mathbb{R}^n} \eta_\varepsilon(x - y) v(y) \, dy = \int_{B(x,\varepsilon)} \eta_\varepsilon(x - y) v(y) \, dy.$$

Then $v^\varepsilon \in C^\infty(\mathbb{T}^n)$, and $v^\varepsilon \to v$ uniformly in $\mathbb{T}^n$ as $\varepsilon \to 0$. As $H \in C^2(\mathbb{T}^n \times \mathbb{R}^n)$, by repeating the proof of Theorem 3.8, we infer that v^ε satisfies

$$H(y, p + Dv^\varepsilon(y)) \le \overline{H}(p) + C\varepsilon \qquad \text{in } \mathbb{T}^n.$$

We now perform a similar computation as the heuristic one above

$$\begin{aligned}
&\int_{-T}^{0} \left(L(\gamma(t),\dot\gamma(t)) + \overline{H}(p)\right)\, dt\\
\geq\ &\int_{-T}^{0} \left(L(\gamma(t),\dot\gamma(t)) + H(\gamma(t), p + Dv^{\varepsilon}(\gamma(t))) - C\varepsilon\right)\, dt\\
\geq\ &-CT\varepsilon + \int_{-T}^{0} \dot\gamma(t)\cdot (p + Dv^{\varepsilon}(\gamma(t)))\, dt\\
=\ &-CT\varepsilon + p\cdot(\gamma(0)-\gamma(-T)) + v^{\varepsilon}(\gamma(0)) - v^{\varepsilon}(\gamma(-T)).
\end{aligned}$$

Let $\varepsilon \to 0$ in the above to conclude. □

3.5. Rotation vectors

Theorem 3.15. *Assume* (3.12). *Fix $x, p \in \mathbb{R}^n$. Let $v \in \mathrm{Lip}\,(\mathbb{T}^n)$ be a solution to* (3.7). *Let ξ be a backward characteristic of v emanating from x. Then, there exist a subsequence $\{t_k\} \to -\infty$ and a vector $q \in D^{-}\overline{H}(p)$ such that*

$$\lim_{k\to\infty} \frac{\xi(t_k)}{t_k} = q \in D^{-}\overline{H}(p) = \partial \overline{H}(p).$$

We say that q is a rotation vector corresponding to the backward characteristic ξ.

Proof. For each $p \in \mathbb{R}^n$, we write v_p to denote a solution to (3.7).

As ξ is a backward characteristic of $v = v_p$ emanating from x, for every $t < 0$,

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

On the other hand, for any $\tilde p \in \mathbb{R}^n$, let $v_{\tilde p} \in \mathrm{Lip}\,(\mathbb{T}^n)$ be a solution to the corresponding cell problem (3.7) with $p = \tilde p$ such that $\min_{\mathbb{T}^n} v_{\tilde p} = 0$. We use Lemma 3.14 to get one-sided control

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

Thus, for $\tilde p \in B(p,1)$,

$$\overline{H}(\tilde p) - \overline{H}(p) \geq (\tilde p - p)\cdot \frac{\xi(t)-\xi(0)}{t} - \frac{C}{|t|}. \tag{3.15}$$

Besides, the fact that $\|\dot\xi\|_{L^\infty((-\infty,0])} \leq C = C(|p|)$ implies

$$\left|\frac{\xi(t)-\xi(0)}{t}\right| \leq C \qquad \text{for all } t<0.$$

Therefore, there exists a sequence $\{t_k\} \to -\infty$ such that $\frac{\xi(t_k)}{t_k} \to q \in \mathbb{R}^n$ as $k \to \infty$ with $|q| \le C$. Plug this into (3.15) to yield

$$\overline{H}(\tilde{p}) - \overline{H}(p) \ge (\tilde{p} - p) \cdot q \qquad \text{for all } \tilde{p} \in B(p,1),$$

which gives that $q \in \partial \overline{H}(p)$.

$\square$

It is unclear whether for different subsequences of $\{t_k\}$, we have different rotation vectors in the limit in case that $\overline{H}$ is not differentiable at p.

3.6. The weak KAM theorem via PDE viewpoint

The following result is often known as the weak KAM theorem in the literature. It is the combination of Theorems 3.1 and 3.10.

Theorem 3.16. *Assume* (3.12). *Fix* $p \in \mathbb{R}^n$. *Then, there exists a Lipschitz viscosity solution* $v \in \mathrm{Lip}\,(\mathbb{T}^n)$ *of the cell problem* (3.7). *Moreover, for each* $x \in \mathbb{R}^n$, *there exists a* C^k *curve* $\xi : (-\infty, 0] \to \mathbb{R}^n$ *such that* $\xi(0) = x$, $\|\dot{\xi}\|_{L^\infty((-\infty,0])} \le C$ *for some* $C = C(|p|) > 0$, *and for all* $t_2 \le t_1 \le 0$,

$$p \cdot \xi(t_1) + v(\xi(t_1)) - p \cdot \xi(t_2) - v(\xi(t_2)) = \int_{t_2}^{t_1} \left(L(\xi(s), \dot{\xi}(s)) + \overline{H}(p) \right) ds.$$

3.7. References

(1) The cell problems were first studied by Lions, Papanicolaou, Varadhan [**LPV**].

(2) For the weak KAM theorem via dynamical viewpoint, see Fathi [**Fat**].

(3) For further analysis of viscosity of Hamilton–Jacobi equations, see Evans [**Eva10**, Chapter 10], Tran [**Tra21**].

Chapter 4

The weak KAM theorem via dynamical system viewpoint

In this chapter, we always consider a given Hamiltonian $H : \mathbb{T}^n \times \mathbb{R}^n \to \mathbb{R}$ that satisfies

$$\begin{cases} H \in C^k(\mathbb{T}^n \times \mathbb{R}^n) \text{ for some } k \geq 2, \\ D^2_{pp}H(y,p) > 0 \text{ for all } (y,p) \in \mathbb{T}^n \times \mathbb{R}^n, \\ \lim_{|p|\to\infty} \min_{y\in\mathbb{T}^n} \frac{H(y,p)}{|p|} = +\infty. \end{cases} \tag{4.1}$$

Let L be the corresponding Lagrangian (the Legendre transform of H). Then, L satisfies

$$\begin{cases} L \in C^k(\mathbb{T}^n \times \mathbb{R}^n), \\ D^2_{vv}L(y,v) > 0 \text{ for all } (y,v) \in \mathbb{T}^n \times \mathbb{R}^n, \\ \lim_{|v|\to\infty} \min_{y\in\mathbb{T}^n} \frac{L(y,v)}{|v|} = +\infty. \end{cases} \tag{4.2}$$

The main object in this chapter is the cell problem at $p = 0$, that is,

$$H(y, Dv(y)) = \overline{H}(0) = c \qquad \text{in } \mathbb{T}^n. \tag{4.3}$$

Here, $c = \overline{H}(0) \in \mathbb{R}$ is the unique constant so that (4.3) has a viscosity solution as discussed in the previous chapter. Sometimes, $c = \overline{H}(0)$ is also called the ergodic constant in the literature. Nevertheless, let us ignore this point for now and deal with (4.3) directly first. We will explain further the significance of this PDE as we proceed.

4.1. Cell problems and Hamiltonian dynamics

Theorem 4.1. *Assume* (4.1). *Assume that* (4.3) *admits a solution* $v \in C^2(\mathbb{T}^n)$. *For each* $x_0 \in \mathbb{T}^n$, *let* $p_0 = Dv(x_0)$, *and we look at the usual Hamiltonian system*

$$\begin{cases} \dot{X}(t) = D_pH(X(t), P(t)), \\ \dot{P}(t) = -D_xH(X(t), P(t)), \\ X(0) = x_0, P(0) = p_0. \end{cases} \tag{4.4}$$

Then,

$$P(t) = Dv(X(t)) \qquad \text{for all } t \in \mathbb{R}.$$

Proof. Here, all vectors are written as column vectors. Consider $(\tilde{X}, \tilde{P})$ such that

$$\begin{cases} \dot{\tilde{X}}(t) = D_pH(\tilde{X}(t), Dv(\tilde{X}(t))), \\ X(0) = x_0, \end{cases} \tag{4.5}$$

and $\tilde{P}(t) = Dv(\tilde{X}(t))$ for all $t \in \mathbb{R}$. We aim at showing that

$$(\tilde{X}, \tilde{P}) = (X, P).$$

To do so, we first calculate that

$$\dot{\tilde{P}}(t) = D^2v(\tilde{X}(t))\dot{\tilde{X}}(t) = D^2v(\tilde{X}(t))D_pH(\tilde{X}(t), Dv(\tilde{X}(t))). \tag{4.6}$$

Besides, we have

$$H(x, Dv(x)) = c \qquad \text{in } \mathbb{T}^n.$$

Differentiate this with respect to x to deduce that

$$D_xH(x, Dv(x)) + D^2v(x)D_pH(x, Dv(x)) = 0. \tag{4.7}$$

Combine (4.6) and (4.7) to get that

$$\dot{\tilde{P}}(t) = -D_xH(\tilde{X}(t), Dv(\tilde{X}(t))) = -D_xH(\tilde{X}(t), \tilde{P}(t)).$$

Thus, $(\tilde{X}, \tilde{P})$ solves (4.4), and hence by the uniqueness of solutions to (4.4), we conclude that $(\tilde{X}, \tilde{P}) = (X, P)$. □

By using the above theorem, we immediately arrive at the following result.

Theorem 4.2. *Assume* (4.1). *Assume that* (4.3) *admits a solution* $v \in C^2(\mathbb{T}^n)$. *Then, for* $(x_0, p_0) = (x_0, Dv(x_0))$ *for some given* $x_0 \in \mathbb{T}^n$, *we have*

$$\phi_t^H(x_0, p_0) = (X(t), P(t)) = (X(t), Du(X(t))) \qquad \text{for all } t \in \mathbb{R}.$$

In particular, the graph

$$\Gamma = \{(x, Du(x)) \, : \, x \in \mathbb{T}^n\} \subset \mathbb{T}^n \times \mathbb{R}^n$$

is invariant under the Hamiltonian flow ϕ_t^H. *Here, by invariance, we mean*

$$\phi_t^H(\Gamma) \subset \Gamma \qquad \text{for all } t \in \mathbb{R}.$$

Remark 4.3. Some important comments for the above two theorems are in order.

First, the assumption that $v \in C^2(\mathbb{T}^n)$ is very restrictive in general. We will see that this cannot hold true in many cases.

Second, one interesting point from the above proof is that if for some reasons that we know v, then in order to solve the Hamiltonian system with $2n$ unknowns, we only need to consider (4.5) with n unknowns, which makes the task simpler.

Finally, for the Hamiltonian system (4.4), we have already shown that $t \mapsto H(X(t), P(t))$ is constant. Therefore, it is natural to consider (4.3) with the fixed energy level $c = H(0) \in \mathbb{R}$ as the Hamiltonian flow keeps the energy conserved anyway. This is one of the reasons why it is rather natural to study cell problems.

As noted by the first point in the above remark, it is more natural to consider solutions to the cell problem (4.3) with lower regularity. We now focus on the situation where (4.3) has C^1 solutions.

Proposition 4.4. *Assume* (4.1). *Let* $v \in C^1(\mathbb{T}^n)$ *be a given function. Then, the following claims are equivalent.*

(i) $H(x, Dv) \le c$ *in* $\mathbb{T}^n$ *for a given constant* $c \in \mathbb{R}$.

(ii) *For every curve* $\gamma \in \mathrm{AC}([a,b], \mathbb{T}^n)$ *for* $a < b$, *we have*

$$v(\gamma(b)) - v(\gamma(a)) \le \int_a^b (L(\gamma(s), \dot{\gamma}(s)) + c)\, ds.$$

We note that a part of this proposition is a weaker version of Lemma 3.14 in the previous chapter.

Proof. We first show "(i) ⇒ (ii)". By the fundamental theorem of calculus and Fenchel's inequality,

$$\begin{aligned} v(\gamma(b)) - v(\gamma(a)) &= \int_a^b Dv(\gamma(s)) \cdot \dot{\gamma}(s)\, ds \\ &\le \int_a^b (H(\gamma(s), Dv(\gamma(s)) + L(\gamma(s), \dot{\gamma}(s)))\, ds \\ &\le \int_a^b (L(\gamma(s), \dot{\gamma}(s)) + c)\, ds. \end{aligned}$$

We next prove "(ii) $\Rightarrow$ (i)". Fix $x \in \mathbb{T}^n$ and a direction $v \in \mathbb{R}^n$. For $t > 0$ small, consider

$$\gamma(s) = x + sv \qquad \text{for } 0 \le s \le t.$$

We have, in light of (ii),

$$\begin{aligned} v(x+tv) - v(x) &= v(\gamma(t)) - v(\gamma(0)) \\ &\le \int_0^t \left(L(\gamma(s), \dot\gamma(s)) + c\right) \, ds = \int_0^t \left(L(x+sv, v) + c\right) \, ds. \end{aligned}$$

Divide both sides of this by t and let $t \to 0$ to imply

$$Dv(x) \cdot v \le L(x,v) + c.$$

Thus,

$$H(x, Dv(x)) = \sup_{v \in \mathbb{R}^n} \left(Dv(x) \cdot v - L(x,v)\right) \le c.$$

$\square$

In the second part of the above proof, we used a general idea to go from the action functional and Dynamic Programming Principle to PDE. We next give a definition of dominated functions in which differentiability is not required.

Definition 4.5 (Dominated functions)**.** Let $u \in C(\mathbb{T}^n)$ and $c \in \mathbb{R}$. We say that u is dominated by $L + c$ on $\mathbb{T}^n$, which we denote by $u \prec L + c$, if for each continuous piecewise C^1 curve $\gamma : [a,b] \to \mathbb{T}^n$, we have

$$u(\gamma(b)) - u(\gamma(a)) \le \int_a^b \left(L(\gamma(s), \dot\gamma(s)) + c\right) \, ds.$$

We also define

$$D^c(\mathbb{T}^n) = \left\{u \in C(\mathbb{T}^n) \, : \, u \prec L + c\right\}.$$

Remark 4.6. In the above definition, by usual approximations, continuous piecewise C^1 curves can be replaced by C^∞ curves or absolutely continuous curves.

Our next important goal is to show that in the case that $v \in C^1(\mathbb{T}^n)$ is a solution to (4.3), then we still have that its graph is invariant under the Hamiltonian flow. This will be done in the next section.

4.2. Invariance under the Hamiltonian and Lagrangian flows

We recall the following points from Theorem 1.15 and Definition 1.16. The map

$$\begin{aligned} \mathcal{L} : \mathbb{T}^n \times \mathbb{R}^n &\to \mathbb{T}^n \times \mathbb{R}^n \\ (x,v) &\mapsto (x,p) = (x, D_v L(x,v)). \end{aligned} \tag{4.8}$$

is a local C^{k-1} diffeomorphism thanks to (4.1). The inverse of $\mathcal{L}$ is $\mathcal{H}$,

$$\begin{aligned} \mathcal{H} : \mathbb{T}^n \times \mathbb{R}^n &\to \mathbb{T}^n \times \mathbb{R}^n \\ (x,p) &\mapsto (x,v) = (x, D_pH(x,p)), \end{aligned}$$

which is also a local C^{k-1} diffeomorphism. In particular, for $u \in C^1(\mathbb{T}^n)$,

$$(x, Du(x)) = \mathcal{L}(x, D_pH(x, Du(x))).$$

We also have that

$$\phi_t^L = \mathcal{L}^{-1} \circ \phi_t^H \circ \mathcal{L}.$$

By using the above identities, the following proposition is straightforward.

Proposition 4.7. *For $u \in C^1(\mathbb{T}^n)$, denote by*

$$\Gamma = \{(x, Du(x)) \,:\, x \in \mathbb{T}^n\}, \qquad \tilde{\Gamma} = \{(x, D_pH(x, Du(x))) \,:\, x \in \mathbb{T}^n\}.$$

Then, Γ is invariant under ϕ_t^H if and only if $\tilde{\Gamma}$ is invariant under ϕ_t^L.

Here is the main result of this section.

Theorem 4.8. *Assume* (4.1). *Assume that* (4.3) *admits a solution $u \in C^1(\mathbb{T}^n)$. Then, the graph*

$$\Gamma = \{(x, Du(x)) \,:\, x \in \mathbb{T}^n\} \subset \mathbb{T}^n \times \mathbb{R}^n$$

is invariant under the Hamiltonian flow ϕ_t^H.

To prove this theorem, we need the following preparatory lemma.

Lemma 4.9. *Assume* (4.1). *Assume that* (4.3) *admits a solution $u \in C^1(\mathbb{T}^n)$. Let $\gamma : [a,b] \to \mathbb{T}^n$ be a solution to*

$$\dot{\gamma}(s) = D_pH(\gamma(s), Du(\gamma(s))) \qquad \text{for } s \in (a,b).$$

Then,

$$u(\gamma(b)) - u(\gamma(a)) = \int_a^b \left(L(\gamma(s), \dot{\gamma}(s)) + c\right) ds.$$

In particular, γ is a minimizer of

$$\min_{\substack{\eta \in \mathrm{AC}\,([a,b]) \\ \eta(a)=\gamma(a),\, \eta(b)=\gamma(b)}} I[\eta], \tag{4.9}$$

and $\gamma \in C^k([a,b])$.

Proof. By the ODE for γ and the properties of the Legendre transform, for $s \in (a, b)$,

$$\begin{aligned}
&L(\gamma(s), \dot{\gamma}(s)) + c \\
&= L(\gamma(s), D_pH(\gamma(s), Du(\gamma(s))) + H(\gamma(s), Du(\gamma(s))) \\
&= D_pH(\gamma(s), Du(\gamma(s))) \cdot Du(\gamma(s)) \\
&= \dot{\gamma}(s) \cdot Du(\gamma(s)) = \frac{d}{ds}(u(\gamma(s))).
\end{aligned}$$

We thus get

$$u(\gamma(b)) - u(\gamma(a)) = \int_a^b (L(\gamma(s), \dot{\gamma}(s)) + c)\, ds.$$

Thanks to Proposition 4.4, γ is a minimizer of (4.9). We then use the regularity theory to yield γ satisfies the corresponding Euler-Lagrange equations and $\gamma \in C^k([a, b])$. □

Proof of Theorem 4.8. By Proposition 4.7, we only need to show that $\tilde{\Gamma}$ is invariant under ϕ_t^L. For

$$(x_0, v_0) = (x_0, D_pH(x_0, Du(x_0))) = (\gamma(0), \dot{\gamma}(0)) \in \tilde{\Gamma},$$

let $\gamma : \mathbb{R} \to \mathbb{T}^n$ be a solution to

$$\dot{\gamma}(s) = D_pH(\gamma(s), Du(\gamma(s))) \qquad \text{for } s \in \mathbb{R}.$$

We use Lemma 4.9 above to yield γ satisfies the corresponding Euler-Lagrange equations and $\gamma \in C^k(\mathbb{R})$. Therefore, for $t \in \mathbb{R}$,

$$\phi_t^L(x_0, v_0) = (\gamma(t), \dot{\gamma}(t)) = (\gamma(t), D_pH(\gamma(t), Du(\gamma(t)))) \in \tilde{\Gamma}.$$

The proof is then complete as $\phi_t^L(\tilde{\Gamma}) \subset \tilde{\Gamma}$. □

Remark 4.10. It is important noting that the existence of γ solving the ODE

$$\dot{\gamma}(s) = D_pH(\gamma(s), Du(\gamma(s))) \qquad \text{for } s \in \mathbb{R}$$

follows from Peano's theorem. Since we only have $u \in C^1(\mathbb{T}^n)$, the vector field $x \mapsto D_pH(x, Du(x))$ is only continuous, and thus, no uniqueness of γ is guaranteed.

The above proofs lead us naturally to the following definition of calibrated curves.

Definition 4.11 (Calibrated curves)**.** Let $I \subset \mathbb{R}$ be an interval and $u \in C(\mathbb{T}^n)$. We say that a continuous, piecewise C^1 curve $\gamma : I \to \mathbb{T}^n$ is (u, L, c)-calibrated if for every $t, t' \in I$ with $t < t'$, we have

$$u(\gamma(t')) - u(\gamma(t)) = \int_t^{t'} (L(\gamma(s), \dot{\gamma}(s)) + c)\, ds.$$

It is clear from the above definition that if $\gamma : I \to \mathbb{T}^n$ is (u, L, c)-calibrated, then $\gamma|_{I'}$ is also (u, L, c)-calibrated for any subinterval $I' \subset I$.

We have already proved the following result, but it is convenient to record it here for later usage.

Theorem 4.12. *Assume* (4.1). *Let $u \in C(\mathbb{T}^n)$ and $c \in \mathbb{R}$ be such that $u \prec L + c$. Let $I \subset \mathbb{R}$ be an interval, and $\gamma : I \to \mathbb{T}^n$ be a (u, L, c)-calibrated curve. Then, $\gamma \in C^k(I)$.*

Proof. Without loss of generality, assume $I = [a, b]$. As γ is (u, L, c)-calibrated, we have

$$u(\gamma(b)) - u(\gamma(a)) = \int_a^b (L(\gamma(s), \dot{\gamma}(s)) + c)\, ds.$$

For every other $\eta \in \mathrm{AC}([a, b], \mathbb{T}^n)$ with $\eta(a) = \gamma(a)$, $\eta(b) = \gamma(b)$, by the definition of $u \prec L + c$,

$$u(\eta(b)) - u(\eta(a)) \leq \int_a^b (L(\eta(s), \dot{\eta}(s)) + c)\, ds.$$

Thus, γ is a minimizer of

$$\min_{\substack{\eta \in \mathrm{AC}([a,b]) \\ \eta(a)=\gamma(a),\, \eta(b)=\gamma(b)}} I[\eta],$$

and hence, $\gamma \in C^k(I)$. □

We have the following characterization of C^1 solutions to the cell problem (4.3).

Proposition 4.13. *Assume* (4.1). *Let $u \in C^1(\mathbb{T}^n)$ and $c \in \mathbb{R}$. The following claims are equivalent.*

(i) *u is a solution to $H(x, Du) = c$ in $\mathbb{T}^n$.*

(ii) *$u \prec L + c$, and for each $x \in \mathbb{T}^n$, we can find $\varepsilon > 0$, and a C^1 curve $\gamma : [-\varepsilon, \varepsilon] \to \mathbb{T}^n$ which is (u, L, c)-calibrated with $\gamma(0) = x$.*

(iii) *$u \prec L + c$, and for each $x \in \mathbb{T}^n$, we can find $\varepsilon > 0$, and a C^1 curve $\gamma : [0, \varepsilon] \to \mathbb{T}^n$ which is (u, L, c)-calibrated with $\gamma(0) = x$.*

(iv) *$u \prec L + c$, and for each $x \in \mathbb{T}^n$, we can find $\varepsilon > 0$, and a C^1 curve $\gamma : [-\varepsilon, 0] \to \mathbb{T}^n$ which is (u, L, c)-calibrated with $\gamma(0) = x$.*

Proof. We first prove "(i) ⇒ (ii)". We already proved $u \prec L + c$ in Proposition 4.4. As $Du(x)$ is continuous, $D_pH(x, Du(x))$ is a continuous vector field. By Peano's theorem, we have short time existence for the following ODE:

$$\begin{cases} \dot{\gamma}(t) = D_pH(\gamma(t), Du(\gamma(t))) & t \in [-\varepsilon, \varepsilon], \\ \gamma(0) = x \end{cases}$$

for a small $\varepsilon > 0$. Then, γ is (u, L, c)-calibrated thanks to Lemma 4.9.

It is obvious that "(ii) $\Rightarrow$ (iii)" and "(ii) $\Rightarrow$ (iv)". To finish off the proof, it is enough to show that "(iv) $\Rightarrow$ (i)". By Proposition 4.4, $H(x, Du) \le c$ in $\mathbb{T}^n$. We only need to show that

$$H(x, Du) \ge c \qquad \text{in } \mathbb{T}^n. \tag{4.10}$$

Fix $x \in \mathbb{T}^n$. By the statement of (iv), there exists a C^1 curve $\gamma : [-\varepsilon, 0] \to \mathbb{T}^n$ which is (u, L, c)-calibrated with $\gamma(0) = x$ for some $\varepsilon > 0$. That is, for $t \in (0, \varepsilon)$,

$$u(\gamma(0)) - u(\gamma(-t)) = \int_{-t}^{0} (L(\gamma(s), \dot\gamma(s)) + c)\ ds.$$

Divide both sides by t and let $t \to 0$ to get

$$Du(x) \cdot \dot\gamma(0) = L(x, \dot\gamma(0)) + c.$$

Thus,

$$H(x, Du(x)) \ge Du(x) \cdot \dot\gamma(0) - L(x, \dot\gamma(0)) \ge c,$$

which confirms (4.10). The proof is complete. □

4.3. Weak KAM solutions

4.3.1. Definition of weak KAM solutions. We first give definitions of weak KAM solutions of negative or positive type.

Definition 4.14 (weak KAM solutions of negative type)**.** Assume (4.1). We say that $u \in C(\mathbb{T}^n)$ is a weak KAM solution of negative type if

- $u \prec L + c$ for some given $c \in \mathbb{R}$;
- for $x \in \mathbb{T}^n$, we can find a (u, L, c)-calibrated C^1 curve $\gamma : (-\infty, 0] \to \mathbb{T}^n$ such that $\gamma(0) = x$.

Let $\mathcal{S}_-$ be the set of all weak KAM solutions of negative type. An element in $\mathcal{S}_-$ is typically denoted as u_-.

Definition 4.15 (weak KAM solutions of positive type)**.** Assume (4.1). We say that $u \in C(\mathbb{T}^n)$ is a weak KAM solution of positive type if

- $u \prec L + c$ for some given $c \in \mathbb{R}$;
- for $x \in \mathbb{T}^n$, we can find a (u, L, c)-calibrated C^1 curve $\gamma : [0, \infty) \to \mathbb{T}^n$ such that $\gamma(0) = x$.

Let $\mathcal{S}_+$ be the set of all weak KAM solutions of negative type. An element in $\mathcal{S}_+$ is typically denoted as u_+.

Remark 4.16. Some important comments concerning the two definitions above in comparison with Proposition 4.13 are in order.

(i) In the two new definitions, we only require that $u \in C(\mathbb{T}^n)$. We do not ask for any differentiability of u, and this is in accordance with the philosophy of viscosity solutions. As a matter of fact, we will see that they are the same, and they represent different facets of the cell problem.

(ii) For the negative and positive type weak KAM solutions, we require the calibrated curves defined on $(-\infty, 0]$ and $[0, \infty)$, respectively. This is a bit different from the intervals $[-\varepsilon, 0]$ and $[0, \varepsilon]$ as in Proposition 4.13.

4.3.2. Characterization of $D^c(\mathbb{T}^n)$. Let us now analyze more about $D^c(\mathbb{T}^n)$ for given $c \in \mathbb{R}$. Recall that

$$D^c(\mathbb{T}^n) = \{u \in C(\mathbb{T}^n) \,:\, u \prec L + c\}.$$

Lemma 4.17. *Assume* (4.1). *Fix $c \in \mathbb{R}$. The following claims hold.*

(i) *If $u \in D^c(\mathbb{T}^n)$, then so is $u + C$ for any $C \in \mathbb{R}$.*

(ii) *$D^c(\mathbb{T}^n)$ is a closed convex subset of $C(\mathbb{T}^n)$.*

(iii) *For $u \in D^c(\mathbb{T}^n)$, we have $u \in \mathrm{Lip}\,(\mathbb{T}^n)$.*

Proof. Item (i) is obvious.

Let us prove (ii). The closedness of $D^c(\mathbb{T}^n)$ is straightforward from its definition. We thus only need to show that it is convex in $C(\mathbb{T}^n)$. Take $u, v \in D^c(\mathbb{T}^n)$. Then, for $\gamma \in \mathrm{AC}\,([a,b], \mathbb{T}^n)$,

$$\begin{aligned} u(\gamma(b)) - u(\gamma(a)) &\le \int_a^b (L(\gamma, \dot\gamma) + c)\; ds, \\ v(\gamma(b)) - v(\gamma(a)) &\le \int_a^b (L(\gamma, \dot\gamma) + c)\; ds. \end{aligned}$$

Then, for $w = ru + (1-r)v$ for $r \in [0,1]$ given, it is clear that

$$w(\gamma(b)) - w(\gamma(a)) \le \int_a^b (L(\gamma, \dot\gamma) + c)\; ds$$

Hence, $w = ru + (1-r)v \in D^c(\mathbb{T}^n)$.

We now prove (iii). We connect any two distinct points $y, z \in \mathbb{T}^n$ by a line segment of unit speed

$$\gamma(s) = y + s\frac{z-y}{|z-y|} \qquad \text{for } 0 \le s \le |z-y|.$$

By definition,

$$\begin{aligned}
u(z)-u(y) &= u(\gamma(|z-y|))-u(\gamma(0))\\
&\le \int_0^{|z-y|}\left(L\left(y+s\frac{z-y}{|z-y|},\frac{z-y}{|z-y|}\right)+c\right)ds\\
&\le \left(\max_{\mathbb{T}^n\times\overline{B}(0,1)} L+c\right)|z-y|\le C|z-y|.
\end{aligned}$$

By a symmetric argument, we conclude the proof. □

By the Arzelà-Ascoli theorem, we immediately deduce the following corollary.

Corollary 4.18. *Assume* (4.1). *Fix $c\in\mathbb{R}$ and $x_0\in\mathbb{T}^n$. Then, the set $\{u-u(x_0)\,:\,u\in D^c(\mathbb{T}^n)\}$ is compact in $C(\mathbb{T}^n)$.*

Let us now find a full characterization of $D^c(\mathbb{T}^n)$.

Lemma 4.19. *Assume* (4.1). *Fix $c\in\mathbb{R}$. Let $u\in D^c(\mathbb{T}^n)$. Then, u is Lipschitz and is differentiable almost everywhere in $\mathbb{T}^n$, and at points of differentiability of u,*

$$H(x,Du(x))\le c \qquad \textit{for a.e. } x\in\mathbb{T}^n.$$

Proof. Thanks to Lemma 4.17(iii), u is Lipschitz. By Rademacher's theorem, u is differentiable almost everywhere in $\mathbb{T}^n$. Pick $x\in\mathbb{T}^n$ to be a point of differentiability of u. Fix $v\in\mathbb{R}^n$ and for $t>0$ small, denote by

$$\gamma(s)=x+sv \qquad \text{for } 0\le s\le t.$$

Then,

$$u(x+tv)-u(x)=u(\gamma(t))-u(\gamma(0))\le\int_0^t(L(\gamma,\dot\gamma)+c)\,ds.$$

Divide both sides by t and let $t\to0$ to yield that

$$Du(x)\cdot v\le L(x,v)+c.$$

Hence,

$$H(x,Du(x))=\sup_{v\in\mathbb{R}^n}(Du(x)\cdot v-L(x,v))\le c.$$

□

We show that the converse of Lemma 4.19 also holds.

Lemma 4.20. *Assume* (4.1). *Fix $c\in\mathbb{R}$. Let $u\in\mathrm{Lip}\,(\mathbb{T}^n)$ be such that*

$$H(x,Du(x))\le c \qquad \textit{for a.e. } x\in\mathbb{T}^n.$$

Then, $u\in D^c(\mathbb{T}^n)$.

Proof. Take η to be the standard mollifier, that is,

$$\eta \in C_c^\infty(\mathbb{R}^n, [0,\infty)), \qquad \operatorname{supp}(\eta) \subset B(0,1), \qquad \int_{\mathbb{R}^n} \eta(x)\,dx = 1.$$

For $\varepsilon > 0$, denote by $\eta_\varepsilon(x) = \varepsilon^{-n}\eta\left(\frac{x}{\varepsilon}\right)$ for all $x \in \mathbb{R}^n$. Set, for $x \in \mathbb{R}^n$,

$$u^\varepsilon(x) = (\eta_\varepsilon \star u)(x) = \int_{\mathbb{R}^n} \eta_\varepsilon(x-y)u(y)\,dy = \int_{B(x,\varepsilon)} \eta_\varepsilon(x-y)u(y)\,dy.$$

Then $u^\varepsilon \in C^\infty(\mathbb{T}^n)$, and $u^\varepsilon \to u$ uniformly in $\mathbb{T}^n$ as $\varepsilon \to 0$. As $H \in C^2(\mathbb{T}^n \times \mathbb{R}^n)$, by repeating the proof of Theorem 3.8, we infer that

$$H(y, Du^\varepsilon(y)) \le c + C\varepsilon \qquad \text{in } \mathbb{T}^n.$$

Then, for $\gamma \in \mathrm{AC}([a,b],\mathbb{T}^n)$,

$$\begin{aligned}
&\int_a^b \left(L(\gamma(t),\dot\gamma(t)) + c\right)\,dt \\
\ge\ &\int_a^b \left(L(\gamma(t),\dot\gamma(t)) + H(\gamma(t), Du^\varepsilon(\gamma(t))) - C\varepsilon\right)\,dt \\
\ge\ &-C(b-a)\varepsilon + \int_a^b \dot\gamma(t)\cdot Du^\varepsilon(\gamma(t))\,dt \\
=\ &-C(b-a)\varepsilon + u^\varepsilon(\gamma(b)) - u^\varepsilon(\gamma(a)).
\end{aligned}$$

Let $\varepsilon \to 0$ in the above to conclude. □

By combining Lemmas 4.19 and 4.20, we have a clear characterization of $D^c(\mathbb{T}^n)$ as follows.

Theorem 4.21 (Characterization of $D^c(\mathbb{T}^n)$)**.** *Assume* (4.1)*. Fix $c \in \mathbb{R}$. Then,*

$$D^c(\mathbb{T}^n) = \{u \in \mathrm{Lip}(\mathbb{T}^n) \,:\, H(x,Du) \le c \text{ a.e. in } \mathbb{T}^n\}.$$

4.3.3. Mañé's critical value. We now define Mañé's critical value. We will see later that this is exactly the same as the effective Hamiltonian (or ergodic constant) at 0.

Definition 4.22 (Mañé's critical value)**.** Assume (4.1). Define Mañé's critical value as

$$c[0] = \inf\{c \in \mathbb{R} \,:\, \text{there exists } u \in \mathrm{Lip}(\mathbb{T}^n) \text{ s.t. } H(x,Du) \le c \text{ a.e. in } \mathbb{T}^n\}.$$

Equivalently,

$$c[0] = \inf_{u \in \mathrm{Lip}(\mathbb{T}^n)} \operatorname*{ess\,sup}_{x \in \mathbb{T}^n} H(x, Du(x)).$$

Of course, by Theorem 4.21, we can also write

$$c[0] = \inf\{c \in \mathbb{R} \,:\, \text{there exists } u \in C(\mathbb{T}^n) \text{ s.t. } u \in D^c(\mathbb{T}^n)\}.$$

Lemma 4.23. *Assume* (4.1). *Let* $c[0]$ *be Mañé's critical value. Then,*

$$\min_{(x,p)\in\mathbb{T}^n\times\mathbb{R}^n} H(x,p) \le c[0] \le \max_{x\in\mathbb{T}^n} H(x,0).$$

Proof. On the one hand, for $\varphi \equiv 0$, we have

$$H(x, D\varphi(x)) = H(x,0) \le \max_{x\in\mathbb{T}^n} H(x,0),$$

which gives $c[0] \le \max_{x\in\mathbb{T}^n} H(x,0)$.

On the other hand, for any $u \in \mathrm{Lip}\,(\mathbb{T}^n)$ such that $H(x,Du) \le c$ a.e. in $\mathbb{T}^n$, we see that

$$\min_{(x,p)\in\mathbb{T}^n\times\mathbb{R}^n} H(x,p) \le H(x,Du(x)) \le c \qquad \text{for a.e. } x \in \mathbb{T}^n,$$

which finishes the proof. $\square$

Theorem 4.24. *Assume* (4.1). *Let* $c[0]$ *be Mañé's critical value. Then, there exists* $u \in \mathrm{Lip}\,(\mathbb{T}^n)$ *such that* $u \prec L + c[0]$, *or in other words,*

$$H(x, Du(x)) \le c[0] \qquad \text{a.e. in } \mathbb{T}^n.$$

Proof. By the definition of $c[0]$, we can find $\{c_k\} \subset \mathbb{R}$, $\{u_k\} \subset \mathrm{Lip}\,(\mathbb{T}^n)$ such that $\lim_{k\to\infty} c_k = c[0]$, and $u_k \prec L + c_k$. For $k \in \mathbb{N}$, denote by

$$\tilde{u}_k(x) = u_k(x) - u_k(0) \qquad \text{for } x \in \mathbb{T}^n.$$

Then, by the coercivity of H in p, there exists $C > 0$ independent of k such that

$$\|\tilde{u}_k\|_{L^\infty(\mathbb{T}^n)} + \|D\tilde{u}_k\|_{L^\infty(\mathbb{T}^n)} \le (\sqrt{n}+1)\|D\tilde{u}_k\|_{L^\infty(\mathbb{T}^n)} \le C.$$

By the Arzelà-Ascoli theorem, there exists a subsequence $\{\tilde{u}_{k_j}\}$ of $\{\tilde{u}_k\}$ such that $\tilde{u}_{k_j} \to u \in C(\mathbb{T}^n)$ uniformly on $\mathbb{T}^n$. It is immediate that $\|Du\|_{L^\infty(\mathbb{T}^n)} \le C$, and thus, $u \in \mathrm{Lip}\,(\mathbb{T}^n)$.

We finally show that $u \prec L + c[0]$ by using the usual stability idea. For $\gamma \in \mathrm{AC}\,([a,b], \mathbb{T}^n)$,

$$u_{k_j}(\gamma(b)) - u_{k_j}(\gamma(a)) \le \int_a^b \left(L(\gamma,\dot\gamma) + c_k\right)\, ds,$$

which is equivalent to

$$\tilde{u}_{k_j}(\gamma(b)) - \tilde{u}_{k_j}(\gamma(a)) \le \int_a^b \left(L(\gamma,\dot\gamma) + c_k\right)\, ds.$$

Let $j \to \infty$ to conclude. $\square$

We present the inf-max representation formula for $c[0]$ (exactly the same to that of $\overline{H}(0)$ in Theorem 3.8), which gives immediately that $c[0] = \overline{H}(0)$.

Theorem 4.25. *Assume* (4.1). *Let* $c[0]$ *be Mañé's critical value. Then,*

$$c[0] = \inf_{\phi \in C^1(\mathbb{T}^n)} \max_{y \in \mathbb{T}^n} H(y, D\phi(y)).$$

Proof. For each $\phi \in C^1(\mathbb{T}^n)$, denote by

$$c_\phi = \max_{y \in \mathbb{T}^n} H(y, D\phi(y)).$$

By the definition of $c[0]$,

$$\inf_{\phi \in C^1(\mathbb{T}^n)} \max_{y \in \mathbb{T}^n} H(y, D\phi(y)) = \inf_{\phi \in C^1(\mathbb{T}^n)} c_\phi \geq c[0].$$

We now prove the reverse inequality. Thanks to Theorem 4.24, there exists $u \in \mathrm{Lip}\,(\mathbb{T}^n)$ such that

$$H(y, Du(y)) \leq c[0] \qquad \text{for a.e. } y \in \mathbb{T}^n. \tag{4.11}$$

We now use convolutions with standard mollifiers to smooth u up. Take η to be a standard mollifier, that is,

$$\eta \in C_c^\infty(\mathbb{R}^n, [0, \infty)), \qquad \mathrm{supp}(\eta) \subset B(0,1), \qquad \int_{\mathbb{R}^n} \eta(x)\, dx = 1.$$

For $\varepsilon > 0$, denote by $\eta_\varepsilon(x) = \varepsilon^{-n}\eta\left(\frac{x}{\varepsilon}\right)$ for all $x \in \mathbb{R}^n$. For $x \in \mathbb{R}^n$, denote by

$$u^\varepsilon(x) = (\eta_\varepsilon \star u)(x) = \int_{\mathbb{R}^n} \eta_\varepsilon(x-y)u(y)\, dy = \int_{B(x,\varepsilon)} \eta_\varepsilon(x-y)u(y)\, dy.$$

Then $u^\varepsilon \in C^\infty(\mathbb{R}^n)$, u^ε is $\mathbb{Z}^n$-periodic, and $u^\varepsilon \to u$ locally uniformly as $\varepsilon \to 0$. By repeating the proof of Theorem 3.8, we imply that

$$H(y, Du^\varepsilon(y)) \leq c[0] + C\varepsilon \qquad \text{in } \mathbb{T}^n.$$

Let $\varepsilon \to 0$ and use the definition of $c[0]$ to conclude. □

To finish off, we have some stability results for calibrated curves.

Proposition 4.26. *Assume* (4.1). *Let $c \in \mathbb{R}$ and $u \in C(\mathbb{T}^n)$. The following claims hold.*

(i) *If $I = \bigcup_{k\in\mathbb{N}} I_k$ and $\{I_k\}$ is a sequence of nested intervals in $\mathbb{R}$ such that $I_1 \subset I_2 \subset \ldots$, and $\gamma : I \to \mathbb{T}^n$ is such that $\gamma|_{I_k}$ is (u, L, c)-calibrated for all $k \in \mathbb{N}$, then γ is (u, L, c)-calibrated on I.*

(ii) *Let $\{\gamma_k\}_{k\in\mathbb{N}} \subset \mathrm{AC}\,([a,b], \mathbb{T}^n)$ be such that $\gamma_k \to \gamma$ in C^1 topology. If γ_k is (u, L, c)-calibrated for all $k \in \mathbb{N}$, then γ is (u, L, c)-calibrated.*

4.3.4. Correspondence between $c[0]$ and calibrated curves. We have the following simple but important correspondence between $c[0]$ and calibrated curves.

Theorem 4.27. *Assume* (4.1). *Let $c \in \mathbb{R}$ and $u \in C(\mathbb{T}^n)$ be such that $u \prec L + c$. Let $\gamma : I \to \mathbb{T}^n$ be (u, L, c)-calibrated where $I \subset \mathbb{R}$ is an interval. If I is of infinite length, then $c = c[0]$.*

Proof. Surely, $c \geq c[0]$.

Assume by contradiction that $c > c[0]$ and I is of infinite length. Pick $u_0 \in C(\mathbb{T}^n)$ such that $u_0 \prec L + c[0]$ in light of Theorem 4.24. For $[a, b] \subset I$, we have

$$\begin{aligned} u(\gamma(b)) - u(\gamma(a)) &= \int_a^b L(\gamma, \dot{\gamma})\, ds + c(b - a), \\ u_0(\gamma(b)) - u_0(\gamma(a)) &\leq \int_a^b L(\gamma, \dot{\gamma})\, ds + c_0(b - a). \end{aligned}$$

Combine these two relations to yield

$$\begin{aligned} (c - c_0)(b - a) &\leq [u(\gamma(b)) - u(\gamma(a))] - [u_0(\gamma(b)) - u_0(\gamma(a))] \\ &\leq \sqrt{n} \left(\|Du\|_{L^\infty(\mathbb{T}^n)} + \|Du_0\|_{L^\infty(\mathbb{T}^n)} \right) \leq C. \end{aligned}$$

As I is of infinite length, we let $b - a \to \infty$ to deduce a contradiction. $\square$

Corollary 4.28. *Assume* (4.1). *To have negative or positive type weak KAM solutions, we must have $c = c[0]$.*

We proceed to study further relations between $u \prec L + c$ and calibrated curves.

Theorem 4.29. *Assume* (4.1). *Let $c \in \mathbb{R}$ and $u \in C(\mathbb{T}^n)$ be such that $u \prec L + c$. Let $\gamma : [a, b] \to \mathbb{T}^n$ be a (u, L, c)-calibrated curve. Then, the following properties hold.*

(i) *If u is differentiable at $\gamma(t)$ for $t \in [a, b]$, then*

$$\begin{cases} Du(\gamma(t)) = D_v L(\gamma(t), \dot{\gamma}(t)), \\ H(\gamma(t), Du(\gamma(t))) = c. \end{cases} \tag{4.12}$$

(ii) *u is differentiable at $\gamma(t)$ for $t \in (a, b)$.*

Proof. Let us first prove (i). Assume $t < b$. In case $t = b$, we use $t' < t$ in the following argument.

Take $t' \in (t, b)$. By the hypothesis,

$$u(\gamma(t')) - u(\gamma(t)) = \int_t^{t'} \left(L(\gamma, \dot{\gamma}) + c \right)\, ds.$$

Divide both sides by $t'-t$ and let $t' \to t$ to imply

$$Du(\gamma(t)) \cdot \dot{\gamma}(t) = L(\gamma(t), \dot{\gamma}(t)) + c,$$

which gives further that

$$H(\gamma(t), Du(\gamma(t))) \geq Du(\gamma(t)) \cdot \dot{\gamma}(t) - L(\gamma(t), \dot{\gamma}(t)) \geq c.$$

On the other hand, as $u \prec L + c$, and u is differentiable at $\gamma(t)$,

$$H(\gamma(t), Du(\gamma(t))) \leq c.$$

We deduce that equality in the above must happen, and therefore,

$$\begin{cases} Du(\gamma(t)) = D_v L(\gamma(t), \dot{\gamma}(t)), \\ H(\gamma(t), Du(\gamma(t))) = c. \end{cases}$$

We now prove (ii). Let $x = \gamma(t)$ for $t \in (a, b)$ fixed. For each $y \in \mathbb{T}^n$, denote by

$$\gamma_y(s) = \gamma(s) + \frac{s-a}{t-a}(y - x) \qquad \text{for } a \leq s \leq t.$$

Then, $\gamma_y(a) = \gamma(a)$, and $\gamma_y(t) = y$. As $u \prec L + c$, we have

$$u(\gamma_y(t)) - u(\gamma_y(a)) \leq \int_a^t (L(\gamma_y, \dot{\gamma}_y) + c) \, ds,$$

which gives

$$u(y) \leq \int_a^t (L(\gamma_y, \dot{\gamma}_y) + c) \, ds + u(\gamma(a)).$$

Define, for $y \in \mathbb{T}^n$,

$$\begin{aligned} \psi^+(y) &= \int_a^t (L(\gamma_y, \dot{\gamma}_y) + c) \, ds + u(\gamma(a)) \\ &= \int_a^t \left(L\left(\gamma(s) + \frac{s-a}{t-a}(y-x), \dot{\gamma}(s) + \frac{y-x}{t-a} \right) + c \right) ds + u(\gamma(a)). \end{aligned}$$

Clearly, $\psi^+ \in C^k$, $\psi^+ \geq u$, and $\psi^+(x) = u(x)$. Geometrically, ψ^+ is a C^k function that touches u from above at x.

We now design in a similar way a C^k function that touches u from below at x. For each $y \in \mathbb{T}^n$, set

$$\eta_y(s) = \gamma(s) + \frac{b-s}{b-t}(y - x) \qquad \text{for } t \leq s \leq b.$$

Then, $\eta_y(t) = y$, and $\eta_y(b) = \gamma(b)$. As $u \prec L + c$, we deduce

$$u(\eta_y(b)) - u(\eta_y(t)) \leq \int_t^b (L(\eta_y, \dot{\eta}_y) + c) \, ds,$$

which yields

$$u(y) \geq -\int_t^b (L(\eta_y, \dot{\eta}_y) + c) \, ds + u(\gamma(b)).$$

Set, for $y \in \mathbb{T}^n$,

$$\begin{aligned}\psi^-(y) &= -\int_t^b (L(\eta_y, \dot\eta_y) + c)\, ds + u(\gamma(b)) \\ &= -\int_t^b \left(L\left(\gamma(s) + \frac{b-s}{b-t}(y-x), \dot\gamma(s) - \frac{y-x}{b-t}\right) + c\right) ds + u(\gamma(b)).\end{aligned}$$

Then, $\psi^- \in C^k$, $\psi^- \le u$, and $\psi^-(x) = u(x)$, which means that ψ^- is a C^k function that touches u from below at x. Hence, ψ^- touches ψ^+ from below at x too. Thus, $D\psi^+(x) = D\psi^-(x)$. By using the definition of differentiability, we see that u is also differentiable at x, and

$$Du(x) = D\psi^+(x) = D\psi^-(x).$$

□

Remark 4.30. The method in the proof above for (ii) is rather important and natural, which is essentially like the variational method. Basically, we use the two family of variations $\{\gamma_y\}_{y\in\mathbb{T}^n}$ of $\gamma|_{[a,t]}$ and $\{\eta_y\}_{y\in\mathbb{T}^n}$ of $\gamma|_{[t,b]}$ to read off information. To get the desired result, we need to have variations from both sides, and it is therefore very important that $t \in (a,b)$.

The differentiability of u might fail at the endpoints $\gamma(a)$ and $\gamma(b)$ of the given calibrated curve in general.

4.3.5. Minimal actions for a given time.

Definition 4.31. For given $x, y \in \mathbb{T}^n$ and $t > 0$, denote by

$$h_t(x,y) = \inf_{\substack{\gamma\in \mathrm{AC}([0,t],\mathbb{T}^n)\\ \gamma(0)=x,\gamma(t)=y}} \int_0^t L(\gamma(s), \dot\gamma(s))\, ds. \tag{4.13}$$

Basically, $h_t(x,y)$ is the minimal cost it takes to travel from x to y in a given fixed amount of time t corresponding to the given Lagrangian L.

We have been dealing with $h_t(x,y)$ all the time up to now, and it is important to summarize things and make them more systematically in this subsection.

Proposition 4.32 (Important properties of h_t)**.** *Assume* (4.1). *We have the following properties of* h_t.

(i) *For* $x, y \in \mathbb{T}^n$ *and* $t > 0$,

$$h_t(x,y) \ge t \min_{(x,v)\in\mathbb{T}^n\times\mathbb{R}^n} L(x,v).$$

(ii) *For* $x, y \in \mathbb{T}^n$ *and* $t, t' > 0$,

$$h_{t+t'}(x,y) = \inf_{z\in\mathbb{T}^n} \left(h_t(x,z) + h_{t'}(z,y)\right).$$

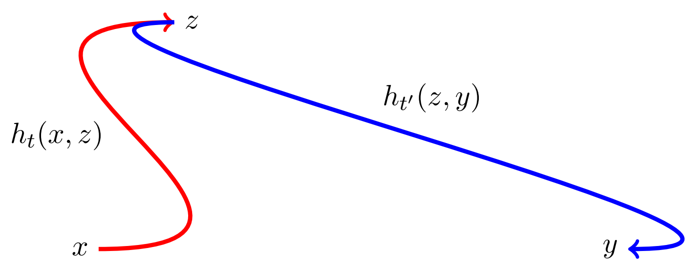

Figure 1. Dynamic Programming Principle for $h_{t+t'}(x,y)$

(iii) *For $u \in C(\mathbb{T}^n)$ and $c \in \mathbb{R}$ such that $u \prec L + c$, then*

$$u(y) - u(x) \le h_t(x,y) + ct \qquad \text{for all } x, y \in \mathbb{T}^n, t > 0.$$

(iv) *For $x, y \in \mathbb{T}^n$ and $t > 0$, there exists an extremal curve $\gamma \in C^k([0,t])$ with $\gamma(0) = x$, $\gamma(t) = y$ such that*

$$h_t(x,y) = \int_0^t L(\gamma(s), \dot\gamma(s))\, ds.$$

Proof. It is straightforward to have item (i). Thanks to Theorems 2.21 and 2.23, we obtain item (iv). For item (iii), we note that for any $\gamma \in \mathrm{AC}([0,t], \mathbb{T}^n)$ with $\gamma(0) = x$, $\gamma(t) = y$,

$$u(y) - u(x) \le \int_0^t L(\gamma, \dot\gamma)\, ds + ct.$$

Take infimum over all such admissible γ to conclude.

Let us now prove (ii), which is basically the Dynamic Programming Principle for h_t. Firstly, for any $\gamma \in \mathrm{AC}([0, t+t']), \mathbb{T}^n)$ with $\gamma(0) = x$, $\gamma(t+t') = y$, we have

$$\begin{aligned} \int_0^{t+t'} L(\gamma, \dot\gamma)\, ds &= \int_0^t L(\gamma, \dot\gamma)\, ds + \int_t^{t+t'} L(\gamma, \dot\gamma)\, ds \\ &\ge h_t(x, \gamma(t)) + h_{t'}(\gamma(t), y) \\ &\ge \inf_{z \in \mathbb{T}^n} \left(h_t(x,z) + h_{t'}(z,y) \right). \end{aligned}$$

Take infimum over all admissible γ to yield

$$h_{t+t'}(x,y) \ge \inf_{z \in \mathbb{T}^n} \left(h_t(x,z) + h_{t'}(z,y) \right).$$

Secondly, for any $\alpha \in \mathrm{AC}([0,t], \mathbb{T}^n)$ with $\alpha(0) = x$, $\alpha(t) = z$, and $\beta \in \mathrm{AC}([0,t'], \mathbb{T}^n)$ with $\beta(0) = z$, $\beta(t') = y$, we define

$$\gamma(s) = \begin{cases} \alpha(s) & \text{for } 0 \le s \le t, \\ \beta(s-t) & \text{for } t \le s \le t + t'. \end{cases}$$

Then, $\gamma \in \mathrm{AC}([0,t+t']),\mathbb{T}^n)$ with $\gamma(0)=x$, $\gamma(t+t')=y$. We see that

$$h_{t+t'}(x,y) \le \int_0^{t+t'} L(\gamma,\dot\gamma)\,ds = \int_0^t L(\alpha,\dot\alpha)\,ds + \int_0^{t'} L(\beta,\dot\beta)\,ds.$$

Take infimum over all possible choices of α and β respectively to conclude. $\square$

Let us analyze more about $h_t(x,y)$.

Lemma 4.33. *Assume* (4.1). *For each $t>0$, there exists $C=C(t)>0$ such that*

$$h_t(x,y) \le C(t) \qquad \text{for all } x,y\in\mathbb{T}^n.$$

Besides, for each $\sigma>0$, there exists $K=K(\sigma)>0$ such that, if $t\ge\sigma$, then for every minimizer γ of $h_t(x,y)$ for $x,y\in\mathbb{T}^n$,

$$|\dot\gamma(s)| \le K(\sigma) \qquad \text{for all } s\in[0,t]. \tag{4.14}$$

Proof. Consider a constant speed line segment connecting x and y

$$\eta(s) = x + s\frac{y-x}{t} \qquad \text{for } 0\le s\le t.$$

Then, it is clear that

$$h_t(x,y) \le \int_0^t L(\eta(s),\dot\eta(s))\,ds \le t \max_{\substack{x\in\mathbb{T}^n\\ |v|\le t^{-1}\sqrt{n}}} |L(x,v)|.$$

We thus can choose

$$C=C(t) = t \max_{\substack{x\in\mathbb{T}^n\\ |v|\le t^{-1}\sqrt{n}}} |L(x,v)|.$$

Let us now prove the second part of the lemma. Let γ be a minimizer of $h_t(x,y)$ for $x,y\in\mathbb{T}^n$ and $t\ge\sigma$. By the mean value theorem, there exists $t_0\in(0,t)$ such that

$$L(\gamma(t_0),\dot\gamma(t_0)) \le \max_{\substack{x\in\mathbb{T}^n\\ |v|\le t^{-1}\sqrt{n}}} |L(x,v)| \le \max_{\substack{x\in\mathbb{T}^n\\ |v|\le \sigma^{-1}\sqrt{n}}} |L(x,v)|.$$

By the superlinearity of L in v, there exists $K=K(\sigma)>0$ such that

$$|\dot\gamma(t_0)| \le K(\sigma) \implies |P(t_0)| = |D_vL(\gamma(t_0),\dot\gamma(t_0))| \le K(\sigma).$$

As $s\mapsto H(X(s),P(s))$ is constant, we see that, for $s\in[0,t]$,

$$H(X(s),P(s)) \le K(\sigma) \implies |P(s)| \le K(\sigma).$$

We also used the superlinearity of H in p in the above. Therefore, for $s\in[0,t]$,

$$|\dot\gamma(s)| = |\dot X(s)| = |D_pH(X(s),P(s))| \le K(\sigma),$$

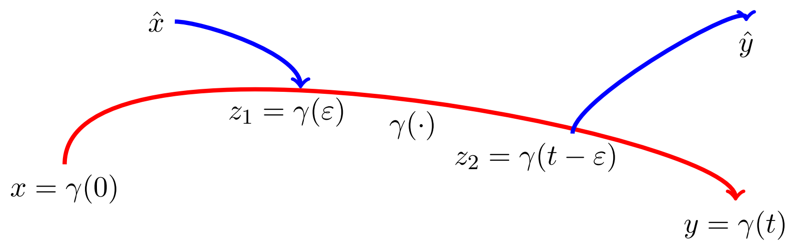

Figure 2. A curve connecting $\hat{x}$ to $\hat{y}$

which completes the proof. Note that $K(\sigma)$ changes from line to line in the above steps. □

We are now ready to prove the following uniform Lipschitz result.

Theorem 4.34. *Assume* (4.1). *For each $\sigma > 0$, there exists $C = C(\sigma) > 0$ such that $h_t : \mathbb{T}^n \times \mathbb{T}^n \to \mathbb{R}$ is Lipschitz with Lipschitz constant at most $C(\sigma)$ for all $t \geq \sigma$.*

Proof. Fix (x, y) and $(\hat{x}, \hat{y})$ in $\mathbb{T}^n \times \mathbb{T}^n$. Take a minimizer path $\gamma : [0, t] \to \mathbb{T}^n$ with $\gamma(0) = x, \gamma(t) = y$, and

$$h_t(x, y) = \int_0^t L(\gamma(s), \dot{\gamma}(s))ds.$$

Fix $\varepsilon > 0$, let $z_1 = \gamma(\varepsilon)$ and $z_2 = \gamma(t-\varepsilon)$, we connect $\hat{x}$ to $\hat{y}$ as following.

Let us define

$$\eta(s) = \begin{cases} \gamma(s) + \frac{\varepsilon - s}{\varepsilon}(\hat{x} - x) & s \in [0, \varepsilon], \\ \gamma(s) & s \in [\varepsilon, t-\varepsilon], \\ \gamma(s) + \frac{s-(t-\varepsilon)}{\varepsilon}(\hat{y} - y) & s \in [t-\varepsilon, t]. \end{cases}$$

Then, η connects $\hat{x}$ to $\hat{y}$ in time t. We have

$$\begin{aligned} & h_t(\hat{x}, \hat{y}) - h_t(x, y) \\ \leq & \int_0^\varepsilon \left[L\left(\gamma(s) + \frac{\varepsilon - s}{\varepsilon}(\hat{x} - x), \dot{\gamma}(s) - \frac{\hat{x} - x}{\varepsilon} \right) - L(\gamma(s), \dot{\gamma}(s)) \right] ds \\ & + \int_{t-\varepsilon}^t \left[L\left(\gamma(s) + \frac{s-(t-\varepsilon)}{\varepsilon}(\hat{y} - y), \dot{\gamma}(s) + \frac{\hat{y} - y}{\varepsilon} \right) - L(\gamma(s), \dot{\gamma}(s)) \right] ds. \end{aligned}$$

It is enough to consider the case $|\hat{x} - x| + |\hat{y} - y| \leq \sigma$. Since $t \geq \sigma$, from Lemma 4.33, we have $|\dot{\gamma}(s)| \leq K(\sigma)$ for all $s \in [0, t]$. Choosing $\varepsilon = \frac{1}{4}\sigma$, we obtain that

$$|\dot{\eta}(s)| \leq K(\sigma) + 4 \qquad \text{for all } s \in [0, t].$$

Thus, there exists $C(\sigma) > 0$ such that, for $x_1, x_2 \in \mathbb{T}^n$ and $|v_1|, |v_2| \leq K(\sigma) + 4$,

$$|L(x_1, v_1) - L(x_2, v_2)| \leq C(\sigma)\Big(|x_1 - x_2| + |v_1 - v_2|\Big).$$

We deduce that

$$h_t(\hat{x}, \hat{y}) - h_t(x, y) \leq C(\sigma)(|\hat{x} - x| + |\hat{y} - y|),$$

and by symmetry, we obtain

$$|h_t(\hat{x}, \hat{y}) - h_t(x, y)| \leq C(\sigma)(|\hat{x} - x| + |\hat{y} - y|).$$

□

4.3.6. The Lax-Oleinik semigroup. Given $g \in C(\mathbb{T}^n)$, we define the Lax-Oleinik semigroup as follows. For $t > 0$,

$$T_t^- g(x) = w(x, t) = \inf_{y \in \mathbb{T}^n} \{g(y) + h_t(y, x)\}.$$

For $t = 0$, set $T_0^- g = g$.

Definition 4.35 (the Lax-Oleinik semigroup)**.** The map $T_t^- : C(\mathbb{T}^n) \to C(\mathbb{T}^n)$ is called the Lax-Oleinik semigroup.

In fact, $w(x, t) = T_t^- g(x)$ is the viscosity solution to the Cauchy problem

$$\begin{cases} w_t + H(x, Dw) = 0 & \text{in } \mathbb{T}^n \times (0, \infty), \\ w(x, 0) = g(x) & \text{on } \mathbb{T}^n. \end{cases}$$

And the Lax-Oleinik semigroup is exactly the optimal control formula for Cauchy problems. We will make everything clear on this aspect later.

For now, we proceed to investigate properties of T_t^-.

Proposition 4.36. *Assume* (4.1)*. The following properties hold.*

(i) *For $g \in C(\mathbb{T}^n)$, $t > 0$, and $x \in \mathbb{T}^n$,*

$$\min_{\mathbb{T}^n} g + t \min_{\mathbb{T}^n \times \mathbb{R}^n} L \leq T_t^- g(x) \leq \min_{\mathbb{T}^n} g + \max_{\mathbb{T}^n \times \mathbb{T}^n} h_t(\cdot, \cdot).$$

(ii) *For fixed $\sigma > 0$, there exists $C(\sigma) > 0$ such that, if $g \in C(\mathbb{T}^n)$ and $t > \sigma$, then $T_t^- g$ is Lipschitz with Lipschitz constant at most $C(\sigma)$.*

Proof. The bounds in item (i) are straightforward.

Item (ii) is quite interesting as although we only start with $g \in C(\mathbb{T}^n)$, T_t^- has a uniform Lipschitz regularizing effect for $t > \sigma$. Fix $x, z \in \mathbb{T}^n$. There exists $\bar{x} \in \mathbb{T}^n$ such that

$$T_t^- g(x) = \min_{y \in \mathbb{T}^n} (h_t(y, x) + g(y)) = h_t(\bar{x}, x) + g(\bar{x}).$$

It is clear that

$$T_t^- g(z) \le h_t(\bar{x}, z) + g(\bar{x}).$$

For $t > \sigma$, we use Theorem 4.34 and the above points to get

$$T_t^- g(z) - T_t^- g(x) \le h_t(\bar{x}, z) - h_t(\bar{x}, x) \le C(\sigma)|z - x|.$$

By a symmetric argument, we imply

$$|T_t^- g(z) - T_t^- g(x)| \le C(\sigma)|z - x|.$$

□

Next, we list further properties of T_t^-.

Proposition 4.37. *Assume* (4.1). *The following properties hold.*

(i) *(Semigroup property) For $t, t' \ge 0$,*

$$T_{t+t'}^- = T_t^- \circ T_{t'}^- = T_{t'}^- \circ T_t^-,$$

and, for $g \in C(\mathbb{T}^n)$, $c \in \mathbb{R}$,

$$T_t^-(g(x) + c) = T_t^-(g(x)) + c.$$

(ii) *(Monotonicity property) If $g, h \in C(\mathbb{T}^n)$ with $g \le h$, then, for $t \ge 0$,*

$$T_t^- g \le T_t^- h.$$

(iii) *(Infimum commutativity) If $\{g_i\}_{i \in I} \subset C(\mathbb{T}^n)$ and $g = \inf_{i \in I} g_i \in C(\mathbb{T}^n)$, then, for $t \ge 0$,*

$$T_t^- g = T_t^- \left(\inf_{i \in I} g_i \right) = \inf_{i \in I} T_t^- g_i.$$

Proof. We first prove (i). For $t, t' \ge 0$, we use Proposition 4.32 to compute

$$\begin{aligned}
T_{t+t'}^- g(x) &= \inf_y \left(g(y) + h_{t+t'}(y, x) \right) \\
&= \inf_y \Big(g(y) + \inf_z \left(h_t(y, z) + h_{t'}(z, x) \right) \Big) \\
&= \inf_{y,z} \left(g(y) + h_t(y, z) + h_{t'}(z, x) \right) \\
&= \inf_z \left(\inf_y \left(g(y) + h_t(y, z) \right) + h_{t'}(z, x) \right) \\
&= \inf_z \big(T_t^- g(z) + h_{t'}(z, x) \big) = T_{t'}^- (T_t^- g)(x).
\end{aligned}$$

It is also clear that for $g \in C(\mathbb{T}^n)$, $c \in \mathbb{R}$,

$$T_t^-(g(x) + c) = T_t^-(g(x)) + c.$$

The monotonicity property (ii) is also straightforward as for $g \le h$ and $t > 0$,

$$T_t^- g(x) = \inf_y \left(g(y) + h_t(y, x) \right) \le \inf_y \left(h(y) + h_t(y, x) \right)) \le T_t^- h(x).$$

Finally, we prove the infimum stability. For $\{g_i\}_{i\in I} \subset C(\mathbb{T}^n)$ with $g = \inf_{i\in I} g_i \in C(\mathbb{T}^n)$ and $t > 0$,

$$\begin{aligned} T_t^- g(x) &= \inf_y \left(g(y) + h_t(y,x)\right) \\ &= \inf_y \left(\inf_{i\in I} g_i(y) + h_t(y,x)\right) \\ &= \inf_{y,i} \left(g_i(y) + h_t(y,x)\right) \\ &= \inf_i \left(\inf_y \left(g_i(y) + h_t(y,x)\right)\right) \\ &= \inf_{i\in I} \left(T_t^- g_i(x)\right). \end{aligned}$$

□

We now prove the contraction property (non-expansiveness property) of T_t^-.

Lemma 4.38 (Non-expansiveness property of T_t^-)**.** *Assume* (4.1)*. For $g_1, g_2 \in C(\mathbb{T}^n)$ and $t \geq 0$,*

$$\|T_t^- g_1 - T_t^- g_2\|_{L^\infty(\mathbb{T}^n)} \leq \|g_1 - g_2\|_{L^\infty(\mathbb{T}^n)}. \tag{4.15}$$

Proof. Let $K = \|g_1 - g_2\|_{L^\infty(\mathbb{T}^n)}$. Then,

$$g_1 - K \leq g_2 \leq g_1 + K.$$

In light of the semigroup property and monotonicity property,

$$T_t^- g_1 - K = T_t^- (g_1 - K) \leq T_t^- g_2 \leq T_t^- (g_1 + K) = T_t^- g_2 + K.$$

Therefore, (4.15) holds true. □

We already showed that for fixed $\sigma > 0$, there exists $C(\sigma) > 0$ such that $T_t^- g$ is Lipschitz with Lipschitz constant at most $C(\sigma)$ for any $g \in C(\mathbb{T}^n)$ and $t > \sigma$. Let us now investigate the continuity of $t \mapsto T_t^- g$.

Lemma 4.39. *Assume* (4.1)*. For a given $g \in C(\mathbb{T}^n)$,*

(i) $\lim_{t\to 0+} T_t^- g = g$;

(ii) *$t \mapsto T_t^- g$ is uniformly continuous.*

Proof. By the non-expansiveness property, it is enough to prove (i) for the case that $g \in \mathrm{Lip}\,(\mathbb{T}^n)$. Assume $\|Dg\|_{L^\infty(\mathbb{T}^n)} \leq K$. Then, for $\gamma(s) = x$ for all $s \in [0,t]$, we see

$$T_t^- g(x) \leq g(x) + \int_0^t L(x,0)\,ds,$$

which gives

$$T_t^- g(x) - g(x) \le tL(x,0). \tag{4.16}$$

Besides, as L is superlinear in v, there exists $C_K > 0$ such that

$$L(x,v) \ge K|v| - C_K \qquad \text{for all } (x,v) \in \mathbb{T}^n \times \mathbb{R}^n.$$

Then, for any $\gamma \in \mathrm{AC}([0,t],\mathbb{R}^n)$ with $\gamma(t) = x$,

$$\begin{aligned}\int_0^t L(\gamma(s),\dot\gamma(s))\,ds &\ge \int_0^t (K|\dot\gamma(s)| - C_K)\,ds \\ &\ge K|\gamma(t)-\gamma(0)| - C_K t = K|x-\gamma(0)| - C_K t.\end{aligned}$$

We hence deduce that

$$\begin{aligned}T_t^- g(x) = \inf_y \left(g(y) + h_t(y,x)\right) &\ge \inf_y \left(g(y) + K|x-y| - C_K t\right) \\ &\ge g(x) - C_K t.\end{aligned} \tag{4.17}$$

Combine (4.16) and (4.17) to conclude that

$$|T_t^- g(x) - g(x)| \le t \max\left\{C_K, \max_{x\in\mathbb{T}^n} L(x,0)\right\}, \tag{4.18}$$

which yields (i).

To prove (ii), we simply use (i), the semigroup property, and the non-expansiveness property. Indeed, for $0 < t < t'$,

$$\begin{aligned}&\|T_{t'}^- g - T_t^- g\|_{L^\infty(\mathbb{T}^n)} = \|T_t^- \circ (T_{t'-t}^- g - g)\|_{L^\infty(\mathbb{T}^n)} \\ &\le \|T_{t'-t}^- g - g\|_{L^\infty(\mathbb{T}^n)} \le (t'-t)\max\left\{C_K, \max_{x\in\mathbb{T}^n} L(x,0)\right\}.\end{aligned}$$

□

Theorem 4.40. *Assume* (4.1)*. Fix $\sigma > 0$. Then the family of functions $\{T_t^- g : g \in C(\mathbb{T}^n)\}$ is equi-Lipschitz on $\mathbb{T}^n \times [\sigma,\infty)$.*

Proof. By Proposition 4.36, for any fixed $t \ge \sigma$, $x \mapsto T_t^- g(x)$ is Lipschitz with Lipchitz constant at most $C(\sigma)$. We then use the proof of Lemma 4.39 above to yield further that, for $t, t' \ge \sigma$ and $g \in C(\mathbb{T}^n)$,

$$\|T_{t'}^- g - T_t^- g\|_{L^\infty(\mathbb{T}^n)} \le |t'-t| \max\left\{C_{C(\sigma)}, \max_{x\in\mathbb{T}^n} L(x,0)\right\}.$$

Summing things up, we deduce, for $g \in C(\mathbb{T}^n)$, $t, t' \ge \sigma$, and $x, y \in \mathbb{T}^n$,

$$\left|T_{t'}^- g(x) - T_t^- g(y)\right| \le \tilde C \left(|x-y| + |t-t'|\right),$$

where

$$\tilde C = \max\left\{C(\sigma), C_{C(\sigma)}, \max_{x\in\mathbb{T}^n} L(x,0)\right\}.$$

□

4.3.7. The weak KAM theorem.

Theorem 4.41 (the weak KAM theorem)**.** *Assume* (4.1). *There exists* $u_- \in C(\mathbb{T}^n)$ *such that*

$$T_t^- u_- + c[0]t = u_- \qquad \text{for all } t \geq 0.$$

Moreover, for each $x \in \mathbb{T}^n$, *there exists a* $(u_-, L, c[0])$*-calibrated curve* $\xi : (-\infty, 0] \to \mathbb{T}^n$ *with* $\xi(0) = x$. *In particular,* u_- *is a weak KAM solution of negative type.*

Proof. We divide the proof into several steps.

Step 1. By Theorem 4.24, there exists $u \in \mathrm{Lip}\,(\mathbb{T}^n)$ such that $u \prec L + c[0]$, or in other words,

$$H(x, Du(x)) \leq c[0] \qquad \text{a.e. in } \mathbb{T}^n.$$

Step 2. Evolve u under T_t^-. We claim that

$$t \mapsto \left(T_t^- u + c[0]t\right) \qquad \text{is nondecreasing.} \tag{4.19}$$

To do this, we first show that, for $t > 0$ and $x \in \mathbb{T}^n$,

$$T_t^- u(x) + c[0]t \geq u(x).$$

Indeed, as $u \prec L + c[0]$, we have, for $\gamma \in \mathrm{AC}\,([0,t], \mathbb{T}^n)$ with $\gamma(t) = x$,

$$u(\gamma(t)) - u(\gamma(0)) \leq \int_0^t L(\gamma, \dot\gamma)\, ds + c[0]t,$$

which means

$$\begin{aligned} u(x) &\leq \inf\left\{\int_0^t L(\gamma, \dot\gamma)\, ds + u(\gamma(0)) \,:\, \gamma \in \mathrm{AC}\,([0,t], \mathbb{T}^n), \gamma(t) = x\right\} + c[0]t \\ &= \inf_{y \in \mathbb{T}^n} \{u(y) + h_t(y,x)\} + c[0]t = T_t^- u(x) + c[0]t. \end{aligned}$$

Then, by the semigroup and monotonicity properties, for $r > 0$,

$$T_r^- u(x) \leq T_r^- \left(T_t^- u(x) + c[0]t\right) = T_{t+r}^- u(x) + c[0]t.$$

Add $c[0]r$ to both sides to yield (4.19).

Step 3. As $u \in \mathrm{Lip}\,(\mathbb{T}^n)$, $T_t^- u(x) + c[0]t$ is globally Lipschitz in $(x,t) \in \mathbb{T}^n \times [0, \infty)$. We now show that there exists $C > 0$ such that

$$T_t^- u(x) + c[0]t \leq C \qquad \text{for all } x \in \mathbb{T}^n, t > 0. \tag{4.20}$$

Assume by contradiction that this is not the case. If for each $t \geq 0$, we can find $x_t \in \mathbb{T}^n$ such that

$$T_t^- u(x_t) + c[0]t \leq u(x_t),$$

then, for every $x \in \mathbb{T}^n$,

$$T_t^- u(x) + c[0]t \leq u(x_t) + C|x - x_t| \leq C,$$

which confirms (4.20) and violates our assumption. Therefore, there exist $\delta, r > 0$ such that

$$T_r^- u(x) + c[0]r \geq u(x) + \delta \qquad \text{for all } x \in \mathbb{T}^n. \tag{4.21}$$

By repeating this multiple number of times, we yield, for all $k \in \mathbb{N}$,

$$T_{kr}^- u(x) + c[0]kr \geq u(x) + k\delta \qquad \text{for all } x \in \mathbb{T}^n.$$

Let $c = c[0] - \delta/r < c[0]$. Then, by the above

$$T_{kr}^- u(x) + ckr \geq u(x) \qquad \text{for all } x \in \mathbb{T}^n. \tag{4.22}$$

For $x \in \mathbb{T}^n$, denote

$$w(x) = \inf_{t \geq 0} \left(T_t^- u(x) + ct\right).$$

Thanks to (4.22), w is well-defined and finite. In fact, by the semigroup and monotonicity properties,

$$w(x) = \inf_{0 \leq t \leq r} \left(T_t^- u(x) + ct\right).$$

Of course, $w \in \mathrm{Lip}\,(\mathbb{T}^n)$. We claim that $w \prec L + c$, which gives a contradiction as $c < c[0]$. Indeed, for $s \geq 0$,

$$\begin{aligned} T_s^-(w + cs) &= T_s^- \left(\inf_{t \geq 0}(T_t^- u(x) + ct) + cs\right) \\ &= \inf_{t \geq 0} \left(T_{s+t}^- u(x) + c(s+t)\right) \geq w. \end{aligned}$$

Thus, $w \leq T_s^-(w + cs)$ for all $s > 0$, which gives $w \prec L + c$. We conclude that (4.20) holds.

Step 4. We use (4.19), (4.20), and the fact that $T_t^- u(x) + c[0]t$ is globally Lipschitz in $(x,t) \in \mathbb{T}^n \times [0,\infty)$ to yield that

$$\left(T_t^- u(x) + c[0]t\right) \to u_-(x) \qquad \text{uniformly on } \mathbb{T}^n \text{ as } t \to \infty \tag{4.23}$$

for some $u_- \in \mathrm{Lip}\,(\mathbb{T}^n)$.

Step 5. We next show that

$$T_t^- u_- + c[0]t = u_- \qquad \text{for all } t \geq 0.$$

This is rather clear as

$$\begin{aligned} T_t^- u_-(x) + c[0]t &= T_t^- \left(\lim_{s \to \infty} \left(T_s^- u(x) + c[0]s\right)\right) + c[0]t \\ &= \lim_{s \to \infty} \left(T_{t+s}^- u(x) + c[0](t+s)\right) = u_-(x). \end{aligned}$$

Step 6. Finally, we show the existence of a $(u_-, L, c[0])$-calibrated curve $\xi : (-\infty, 0] \to \mathbb{T}^n$ with $\xi(0) = x$. The proof of this step is similar to that of Theorem 3.10. We construct ξ iteratively on $[-m, -m+1]$ for $m \in \mathbb{N}$. It is enough for us to give the construction of ξ on $[-1,0]$. There exists $z \in \mathbb{T}^n$ such that

$$u_-(x) = T_1^- u_-(x) + c[0] = \min_{y \in \mathbb{T}^n} (h_1(y,x) + u_-(y)) + c[0]$$
$$= h_1(z,x) + u_-(x) + c[0].$$

By Proposition 4.32, there exists $\eta \in C^k([0,1])$ with $\eta(0) = z$, $\eta(1) = x$, and

$$h_1(z,x) = \int_0^1 L(\eta(s), \dot{\eta}(s))\, ds.$$

Set

$$\xi(s) = \eta(s+1) \qquad \text{for all } s \in [-1,0].$$

The proof is complete. $\square$

We give a second proof of the weak KAM theorem by using Schauder's fixed point theorem.

Second proof of Theorem 4.41. The key point that we use in this proof is Theorem 4.40. As usual, we divide the proof into several steps for clarity.

Step 1. Set $E = C(\mathbb{T}^n)/\mathbb{R} \cdot 1$, that is, we put φ and $\varphi + C$ for $C \in \mathbb{R}$ in the same equivalent class in E for each $\varphi \in C(\mathbb{T}^n)$. For each such $\varphi \in C(\mathbb{T}^n)$, we have $[\varphi] \in E$ with

$$[\varphi] = \{\varphi + C \,:\, C \in \mathbb{R}\},$$

and denote

$$\|[\varphi]\|_E = \inf_{C \in \mathbb{R}} \|\varphi + C\|_{L^\infty(\mathbb{T}^n)}.$$

As $T_t^-(u+C) = T_t^- u + C$, we can think of $T_t^- : E \to E$ as well.

Step 2. For each fixed $\sigma > 0$, we see that $T_\sigma^-(E)$ is equi-Lipschitz in $\mathbb{T}^n$ with Lipschitz constant at most $C(\sigma)$. Therefore, for $[\varphi] \in T_\sigma^-(E)$,

$$\|[\varphi]\|_E \le C(\sigma)\sqrt{n}.$$

By the Arzelà-Ascoli theorem, $T_\sigma^-(E)$ is compact in E. By Schauder's fixed point theorem, there exists $[u_\sigma] \in E$ such that

$$T_\sigma^-[u_\sigma] = [u_\sigma].$$

Then, for any $k \in \mathbb{N}$,

$$T_{k\sigma}^-[u_\sigma] = [u_\sigma]. \tag{4.24}$$

Step 3. For each $j \in \mathbb{N}$, let $[u_j]$ be a fixed point to $T_{2^{-j}}^-$. By (4.24), for $k, j \in \mathbb{N}$,

$$T_{k2^{-j}}^-[u_j] = [u_j].$$

As $j \to \infty$, by the Arzelà-Ascoli theorem, up to passing to a subsequence, we can assume $[u_j] \to [u]$ for some $[u] \in E$. By the continuity of $t \mapsto T_t^-$, we see that, for $t \geq 0$,

$$T_t^-[u] = [u]. \tag{4.25}$$

Step 4. Thanks to (4.25), for each $t > 0$, there exists $c(t) \in \mathbb{R}$ such that

$$T_t^- u = u + c(t).$$

It is clear that $t \mapsto c(t)$ is additive, that is, for $t, s > 0$,

$$c(t+s) = c(t) + c(s).$$

As T_t^- is continuous, so is $c(t)$. Therefore, there exists $c \in \mathbb{R}$ such that

$$c(t) = -ct \qquad \text{for all } t \geq 0.$$

We thus get

$$T_t^- u + ct = u \qquad \text{for all } t \geq 0. \tag{4.26}$$

Step 5. By repeating Step 6 of the first proof, we have the existence of a (u, L, c)-calibrated curve $\xi : (-\infty, 0] \to \mathbb{T}^n$ with $\xi(0) = x$. Thanks to Theorem 4.27, $c = c[0]$.

$\square$

4.4. References

(1) The cell problems were first studied by Lions, Papanicolaou, Varadhan [**LPV**].

(2) For the weak KAM theorem via dynamical viewpoint, see Fathi [**Fat**]. In fact, this chapter is heavily based on [**Fat**]. See also the books of Gomes [**Gom09**], Sorrentino [**Sor15**], Tran [**Tra21**].

(3) There are many excellent survey papers and lecture notes in weak KAM theory: see Evans [**Eva08, Eva04**], Ishii [**Ish**], Kaloshin [**Kal05**], and the references therein.

Chapter 5

Invariant measures

In this chapter, we always consider a given Hamiltonian $H : \mathbb{T}^n \times \mathbb{R}^n \to \mathbb{R}$ that satisfies

$$(5.1)\qquad \begin{cases} H \in C^k(\mathbb{T}^n \times \mathbb{R}^n) \text{ for some } k \geq 2, \\ D^2_{pp}H(y,p) > 0 \text{ for all } (y,p) \in \mathbb{T}^n \times \mathbb{R}^n, \\ \lim_{|p|\to\infty} \min_{y\in\mathbb{T}^n} \frac{H(y,p)}{|p|} = +\infty. \end{cases}$$

Let L be the corresponding Lagrangian (the Legendre transform of H). Then, L satisfies

$$(5.2)\qquad \begin{cases} L \in C^k(\mathbb{T}^n \times \mathbb{R}^n), \\ D^2_{vv}L(y,v) > 0 \text{ for all } (y,v) \in \mathbb{T}^n \times \mathbb{R}^n, \\ \lim_{|v|\to\infty} \min_{y\in\mathbb{T}^n} \frac{L(y,v)}{|v|} = +\infty. \end{cases}$$

The main object in this chapter is still the cell problem at $p = 0$, that is,

$$(5.3)\qquad H(y, Dv(y)) = \overline{H}(0) = c[0] \qquad \text{in } \mathbb{T}^n.$$

Here, $c[0] = \overline{H}(0) \in \mathbb{R}$ is the unique constant so that (4.3) has a viscosity solution as discussed in the previous chapters. Sometimes, $c[0] = \overline{H}(0)$ is also called the ergodic constant in the literature.

We have proved the weak KAM theorem (Theorem 4.41) using the dynamical system viewpoint in the previous chapter. The equivalent form of this weak KAM theorem is Theorem 3.16 from the PDE viewpoint. Basically, the weak KAM theorem asserts that (5.3) has a viscosity solution $u \in \mathrm{Lip}\,(\mathbb{T}^n)$, which is equivalent to the fact that

$$T^-_t u + c[0]t = u \qquad \text{for all } t \geq 0.$$

Here, we write u instead of u_- for clarity. Besides, for each $x \in \mathbb{T}^n$, there exists a $(u, L, c[0])$-calibrated curve $\xi : (-\infty, 0] \to \mathbb{T}^n$ with $\xi(0) = x$. More precisely, for $r < r' \leq 0$, we have

$$u(\xi(r')) - u(\xi(r)) = \int_r^{r'} \Big(L(\xi(s), \dot{\xi}(s)) + c[0]\Big)\, ds.$$

In the PDE language, ξ is also called a backward characteristic of u emanating from x. In particular, u is a weak KAM solution of negative type.

If we view the calibrated curve ξ as a curve in $\mathbb{R}^n$, then ξ has corresponding rotation vectors. By Theorem 3.15, there exist a subsequence $\{t_k\} \to -\infty$ and a vector $q \in \partial \overline{H}(0)$ such that

$$\lim_{k\to\infty} \frac{\xi(t_k)}{t_k} = q. \tag{5.4}$$

If $\overline{H}$ is differentiable at 0, that is, $\partial\overline{H}(0)$ is a singleton, then the above limit holds for the full sequence

$$\lim_{t\to-\infty} \frac{\xi(t)}{t} = D\overline{H}(0).$$

If $\overline{H}$ is not differentiable at 0, then it is not yet clear whether we have different subsequences convergent to different rotation vectors.

The main goal of this chapter is to study further properties of u and ξ. Recall that we proved in Theorem 4.29 that u is differentiable at $\xi(t)$ for all $t \in (-\infty, 0)$, and

$$Du(\xi(t)) = D_v L(\xi(t), \dot{\xi}(t)). \tag{5.5}$$

5.1. Flow invariance

Recall first the Lagrangian flow

$$\begin{cases} \phi_t^L(x, v) = (x(t), \dot{x}(t)) = (x(t), v(t)) & \text{for } t \in \mathbb{R}, \\ (x(0), \dot{x}(0)) = (x(0), v(0)) = (x, v). \end{cases}$$

Here, $x(\cdot)$ solves the Euler-Lagrange equations

$$\frac{d}{dt}\left(D_v L(x(t), \dot{x}(t))\right) = D_x L(x(t), \dot{x}(t)).$$

Definition 5.1 (Flow invariance measures)**.** A Radon probability measure $\mu \in \mathcal{P}(\mathbb{T}^n \times \mathbb{R}^n)$ is said to be flow invariant if for every bounded continuous function $\psi : \mathbb{T}^n \times \mathbb{R}^n \to \mathbb{R}$ and every $t \geq 0$,

$$\int_{\mathbb{T}^n\times\mathbb{R}^n} \psi\left(\phi_t^L(x, v)\right)\, d\mu(x, v) = \int_{\mathbb{T}^n\times\mathbb{R}^n} \psi(x, v)\, d\mu(x, v).$$

We also say that μ is invariant under the Euler-Lagrange flow.

Here is another characterization of $c[0]$.

Theorem 5.2. *Assume* (5.1). *Then,*

$$c[0] = -\inf\left\{\int_{\mathbb{T}^n\times\mathbb{R}^n} L(x,v)\,d\mu(x,v) \,:\, \mu\in\mathcal{P}(\mathbb{T}^n\times\mathbb{R}^n) \text{ is flow invariant}\right\}.$$

Proof. For each $(x,v)\in\mathbb{T}^n\times\mathbb{R}^n$, let $(x(t),\dot{x}(t))=(x(t),v(t))$ be the Euler-Lagrange curve as above. As $u\prec L+c[0]$,

$$u(x)-u(x(-1))\le\int_{-1}^{0} L(x(s),\dot{x}(s))\,ds + c[0].$$

Integrate this with respect to $\mu\in\mathcal{P}(\mathbb{T}^n\times\mathbb{R}^n)$ flow invariant to yield

$$\begin{aligned}
0 &= \int_{\mathbb{T}^n\times\mathbb{R}^n}\left(u(\pi\circ\phi_0^L(x,v)) - u(\pi\circ\phi_{-1}^L(x,v))\right)\,d\mu(x,v)\\
&\le \int_{-1}^{0}\int_{\mathbb{T}^n\times\mathbb{R}^n} L(\phi_s^L(x,v))\,d\mu(x,v)\,ds + c[0]\\
&= \int_{\mathbb{T}^n\times\mathbb{R}^n} L(x,v)\,d\mu(x,v) + c[0].
\end{aligned}$$

Thus,

$$\inf\left\{\int_{\mathbb{T}^n\times\mathbb{R}^n} L(x,v)\,d\mu(x,v) \,:\, \mu\in\mathcal{P}(\mathbb{T}^n\times\mathbb{R}^n) \text{ is flow invariant}\right\}\ge -c[0].$$

We now prove the converse. Fix $x\in\mathbb{T}^n$, and let ξ be a $(u,L,c[0])$-calibrated curve $\xi:(-\infty,0]\to\mathbb{T}^n$ with $\xi(0)=x$. For each $t<0$,

$$u(\xi(0))-u(\xi(t))=\int_t^0\left(L(\xi(s),\dot{\xi}(s))+c[0]\right)\,ds.$$

Define $\mu_t\in\mathcal{P}(\mathbb{T}^n\times\mathbb{R}^n)$ as

$$\langle\mu_t,\psi\rangle=\int_{\mathbb{T}^n\times\mathbb{R}^n}\psi(x,v)\,d\mu_t(x,v)=\frac{1}{|t|}\int_t^0\psi(\xi(s),\dot{\xi}(s))\,ds$$

for all bounded continuous functions ψ. As $\|\dot{\xi}\|_{L^\infty((-\infty,0])}\le C$, we see that

$$\operatorname{supp}(\mu_t)\subset\mathbb{T}^n\times\overline{B}(0,C)\qquad\text{for all } t<0.$$

Then, we have

$$\frac{u(x)-u(\xi(t))}{|t|}=\int_{\mathbb{T}^n\times\mathbb{R}^n} L(x,v)\,d\mu_t(x,v)+c[0]. \tag{5.6}$$

By compactness, there exists a sequence $\{t_k\}\to-\infty$ such that

$$\mu_{t_k}\rightharpoonup\mu\in\mathcal{P}(\mathbb{T}^n\times\overline{B}(0,C))\qquad\text{weakly in the sense of measures.}$$

Let $t = t_k$ and $t_k \to -\infty$ in (5.6) to yield

$$\int_{\mathbb{T}^n\times\mathbb{R}^n} L(x,v)\,d\mu(x,v) = -c[0]. \tag{5.7}$$

To finish the proof, we need to show that μ is flow invariant. Indeed, for any bounded continuous function ψ and $t > 0$,

$$\begin{aligned}
&\int_{\mathbb{T}^n\times\mathbb{R}^n} \psi(\phi_t^L(x,v))\,d\mu(x,v) = \lim_{k\to\infty}\frac{1}{|t_k|}\int_{t_k}^0 \psi\circ\phi_t^L(\xi(s),\dot\xi(s))\,ds\\
&= \lim_{k\to\infty}\frac{1}{|t_k|}\int_{t_k}^0 \psi(\xi(s+t),\dot\xi(s+t))\,ds\\
&= \lim_{k\to\infty}\frac{1}{|t_k|}\int_{t_k}^0 \psi(\xi(s),\dot\xi(s))\,ds\\
&\qquad\qquad + \lim_{k\to\infty}\frac{1}{|t_k|}\left[\int_0^t \psi(\xi(s),\dot\xi(s))\,ds - \int_{t_k}^{t_k+t}\psi(\xi(s),\dot\xi(s))\,ds\right]\\
&= \int_{\mathbb{T}^n\times\mathbb{R}^n} \psi(x,v)\,d\mu(x,v) + \lim_{k\to\infty}\frac{C}{|t_k|}\\
&= \int_{\mathbb{T}^n\times\mathbb{R}^n} \psi(x,v)\,d\mu(x,v).
\end{aligned}$$

□

Remark 5.3. Through the construction in the above proof, we have obtained a minimizer μ to the minimizing (variational) problem

$$\inf\left\{\int_{\mathbb{T}^n\times\mathbb{R}^n} L(x,v)\,d\mu(x,v)\ :\ \mu\in\mathcal{P}(\mathbb{T}^n\times\mathbb{R}^n)\text{ is flow invariant}\right\}. \tag{5.8}$$

Note the similarity between $\mu_{t_k} \rightharpoonup \mu$ weakly in the sense of measures and (5.4).

5.2. Mather's measures

We are ready to define Mather's measures as minimizing measures to the variational problem (5.8).

Definition 5.4 (Mather's measures)**.** If μ is a minimizer of (5.8), then we call μ a Mather measure. Denote the Mather set by

$$\widetilde{\mathcal{M}}_0 = \bigcup_{\substack{\mu\text{ is a}\\ \text{Mather measure}}} \operatorname{supp}(\mu).$$

For $\pi : \mathbb{T}^n\times\mathbb{R}^n \to \mathbb{T}^n$ being the natural projection, that is, $\pi(x,v) = x$ for $(x,v)\in\mathbb{T}^n\times\mathbb{R}^n$, the projected Mather set is defined as

$$\mathcal{M}_0 = \pi\left(\widetilde{\mathcal{M}}_0\right).$$

We have the following property of $\widetilde{\mathcal{M}}_0$.

Lemma 5.5. *Assume* (5.1). *Let* $u \in \mathrm{Lip}\,(\mathbb{T}^n)$ *be a solution to* (5.3). *Pick* $(x,v) \in \widetilde{\mathcal{M}}_0$. *Then, for* $t \le t'$,

$$u\left(\pi \circ \phi^L_{t'}(x,v)\right) - u\left(\pi \circ \phi^L_t(x,v)\right) = \int_t^{t'} \left(L\left(\phi^L_s(x,v)\right) + c[0]\right)\, ds. \tag{5.9}$$

Proof. Pick a Mather measure μ such that $(x,v) \in \mathrm{supp}\,(\mu)$. First of all, it is clear that

$$u\left(\pi \circ \phi^L_{t'}(x,v)\right) - u\left(\pi \circ \phi^L_t(x,v)\right) \le \int_t^{t'} \left(L\left(\phi^L_s(x,v)\right) + c[0]\right)\, ds.$$

Integrate this over $d\mu(x,v)$ to yield

$$\begin{aligned} 0 &= \int_{\mathbb{T}^n\times\mathbb{R}^n} u\circ\pi\, d\mu(x,v) - \int_{\mathbb{T}^n\times\mathbb{R}^n} u\circ\pi\, d\mu(x,v) \\ &= \int_{\mathbb{T}^n\times\mathbb{R}^n} u\left(\pi\circ\phi^L_{t'}(x,v)\right)\, d\mu(x,v) - \int_{\mathbb{T}^n\times\mathbb{R}^n} u\left(\pi\circ\phi^L_t(x,v)\right)\, d\mu(x,v) \\ &\le \int_t^{t'} \int_{\mathbb{T}^n\times\mathbb{R}^n} \left(L\left(\phi^L_s(x,v)\right) + c[0]\right)\, d\mu(x,v)ds = 0. \end{aligned}$$

Thus, the inequality in the above must become an equality. Hence, equality must happen on the support of μ, which means that (5.9) holds. $\square$

Lemma 5.6. *Assume* (5.1). *Let* $u \in \mathrm{Lip}\,(\mathbb{T}^n)$ *be a solution to* (5.3). *Then, for* $(x,v) \in \widetilde{\mathcal{M}}_0$, *$u$ is differentiable at x, and*

$$Du(x) = D_vL(x,v).$$

Moreover,

$$\widetilde{\mathcal{M}}_0 \subset \left\{(x,v) \in \mathbb{T}^n\times\mathbb{R}^n \,:\, H(x, D_vL(x,v)) = c[0]\right\}.$$

In particular, $\widetilde{\mathcal{M}}_0$ is contained in the $c[0]$-level set of H and is compact.

Proof. We note that (5.9) holds for $t < 0 < t'$. By Theorem 4.29, u is differentiable at $x(s)$ for all $s \in \mathbb{R}$, and

$$Du(x(s)) = D_vL(x(s), \dot{x}(s)). \tag{5.10}$$

In particular, u is differentiable at $x = x(0)$, and

$$Du(x) = D_vL(x,v).$$

Since u is differentiable at x, (5.3) holds in the classical sense there, and

$$H(x, Du(x)) = H(x, D_vL(x,v)) = c[0].$$

The proof is complete. $\square$

5.2.1. A uniqueness result. We now have a uniqueness result for weak KAM solutions of negative type.

Theorem 5.7. *Assume* (5.1). *Let u_1, u_2 be two weak KAM solutions of negative type. Assume that $u_1 = u_2$ on $\mathcal{M}_0$. Then, $u_1 = u_2$.*

Proof. Fix $x \in \mathbb{T}^n$. Let $\xi : (-\infty, 0] \to \mathbb{T}^n$ be a $(u_1, L, c[0])$-calibrated curve. Then, for any $t < 0$,

$$u_1(x) - u_1(\xi(t)) = \int_t^0 L(\xi, \dot{\xi})\, ds + |t|c[0],$$

$$u_2(x) - u_2(\xi(t)) \le \int_t^0 L(\xi, \dot{\xi})\, ds + |t|c[0].$$

We infer that, for all $t < 0$,

$$u_2(x) - u_1(x) \le u_2(\xi(t)) - u_1(\xi(t)). \tag{5.11}$$

We use ξ to construct a Mather measure as in the proof of Theorem 5.2. Define $\mu_t \in \mathcal{P}(\mathbb{T}^n \times \mathbb{R}^n)$ as

$$\langle \mu_t, \psi \rangle = \int_{\mathbb{T}^n \times \mathbb{R}^n} \psi(x, v)\, d\mu_t(x, v) = \frac{1}{|t|} \int_t^0 \psi(\xi(s), \dot{\xi}(s))\, ds$$

for all bounded continuous functions ψ. As $\|\dot{\xi}\|_{L^\infty((-\infty,0])} \le C$, we see that

$$\operatorname{supp}(\mu_t) \subset \mathbb{T}^n \times \overline{B}(0, C) \qquad \text{for all } t < 0.$$

By compactness, there exists a sequence $\{t_k\} \to -\infty$ such that

$$\mu_{t_k} \rightharpoonup \mu \in \mathcal{P}(\mathbb{T}^n \times \overline{B}(0, C)) \quad \text{weakly in the sense of measures,}$$

and μ is a Mather measure. We use (5.11) to imply

$$u_2(x) - u_1(x) \le \frac{1}{|t_k|} \int_{t_k}^0 (u_2 - u_1)(\xi(s))\, ds = \int_{\mathbb{T}^n \times \mathbb{R}^n} (u_2 - u_1) \circ \pi\, d\mu_{t_k}(x, v).$$

Let $k \to \infty$ to deduce that

$$u_2(x) - u_1(x) \le \int_{\mathbb{T}^n \times \mathbb{R}^n} (u_2 - u_1) \circ \pi\, d\mu(x, v) = 0.$$

By a symmetric argument, we conclude that $u_1 = u_2$. □

5.2.2. Lipschitz graph theorem. In the following, we obtain the famous Lipschitz graph theorem.

Theorem 5.8. *Assume* (5.1). *Let $u \in \operatorname{Lip}(\mathbb{T}^n)$ be a solution to* (5.3). *There exists $C > 0$ such that, for all $x \in \mathcal{M}_0$ and $h \in \mathbb{R}^n$,*

$$|u(x + h) + u(x - h) - 2u(x)| \le C|h|^2.$$

Proof. Fix $(x, v) \in \widetilde{\mathcal{M}}_0$. For $t \in \mathbb{R}$, write $\phi_t^L(x, v) = (x(t), \dot{x}(t))$. By (5.9),

$$
\begin{aligned}
u(x(1)) - u(x(0)) &= \int_0^1 L(x(s), \dot{x}(s))\, ds + c[0], \\
u(x(0)) - u(x(-1)) &= \int_{-1}^0 L(x(s), \dot{x}(s))\, ds + c[0].
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
u(x(1)) - u(x(0) + h) &\le \int_0^1 L(x(s) + (1-s)h, \dot{x}(s) - h)\, ds + c[0], \\
u(x(1)) - u(x(0) - h) &\le \int_{-1}^0 L(x(s) - (1-s)h, \dot{x}(s) + h)\, ds + c[0].
\end{aligned}
$$

Combine the above relations to imply

$$
\begin{aligned}
&u(x+h) + u(x-h) - 2u(x) \\
\ge &\int_0^1 (2L(x, \dot{x}) - L(x + (1-s)h, \dot{x} - h) - L(x - (1-s)h, \dot{x} + h))\ ds \\
\ge &\ -C|h|^2.
\end{aligned}
$$

We obtain the converse bound in a similar way. Indeed,

$$
\begin{aligned}
u(x(0) + h) - u(x(-1)) &\le \int_{-1}^0 L(x(s) + (1+s)h, \dot{x}(s) + h)\, ds + c[0], \\
u(x(0) - h) - u(x(-1)) &\le \int_{-1}^0 L(x(s) - (1+s)h, \dot{x}(s) - h)\, ds + c[0].
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&u(x+h) + u(x-h) - 2u(x) \\
\le &\int_0^1 (L(x + (1+s)h, \dot{x} + h) - L(x - (1+s)h, \dot{x} - h) - 2L(x, \dot{x}))\ ds \\
\le &\, C|h|^2.
\end{aligned}
$$

□

Theorem 5.9. *Assume* (5.1). *Let* $u \in \mathrm{Lip}\,(\mathbb{T}^n)$ *be a solution to* (5.3). *There exists* $C > 0$ *such that, for all* $x \in \mathcal{M}_0$ *and* $h \in \mathbb{R}^n$,

$$
|u(x+h) - u(x) - Du(x) \cdot h| \le C|h|^2.
$$

Proof. We use essentially the ideas in the proof of Theorem 5.8. Fix $(x, v) \in \widetilde{\mathcal{M}}_0$. Then, $Du(x) = D_v L(x, v)$.

By the first part of the proof of Theorem 5.8 and the Euler-Lagrange equations,

$$
\begin{aligned}
& u(x+h)-u(x)\\
\geq & \int_0^1 \left(L(x,\dot{x})-L(x+(1-s)h,\dot{x}-h)\right)\,ds\\
\geq & \int_0^1 \left(D_xL(x,\dot{x})\cdot(s-1)h+D_vL(x,\dot{x})\cdot h\right)\,ds-C|h|^2\\
= & \int_0^1 \left(\frac{d}{ds}\left(D_vL(x,\dot{x})\right)\cdot(s-1)h+D_vL(x,\dot{x})\cdot h\right)\,ds-C|h|^2\\
= & \int_0^1 \frac{d}{ds}\left(D_vL(x,\dot{x})\cdot(s-1)h\right)\,ds-C|h|^2\\
= & \, D_vL(x(0),\dot{x}(0))\cdot h-C|h|^2=Du(x)\cdot h-C|h|^2.
\end{aligned}
$$

The converse bound can be obtained in a similar way by using the second part of the proof of Theorem 5.8 and the Euler-Lagrange equations, and its proof is hence omitted. □

Hidden in the above two theorems are rather deep properties of the differentiability of u along backward characteristics. Let us record them here. It is always fine to go back in time along the backward characteristics, and hence the second part of the proof of Theorem 5.8 always holds true. This leads to the semiconcavity of u.

Theorem 5.10 (semiconcavity)**.** *Assume* (5.1). *Let $u \in \mathrm{Lip}\,(\mathbb{T}^n)$ be a solution to* (5.3). *There exists $C>0$ such that, for all $x \in \mathbb{T}^n$ and $h \in \mathbb{R}^n$,*

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

Besides, we also have local controls along backward characteristics except the endpoints.

Corollary 5.11. *Assume* (5.1). *Let $u \in \mathrm{Lip}\,(\mathbb{T}^n)$ be a solution to* (5.3). *Let $\gamma : (-\infty,0] \to \mathbb{T}^n$ be a backward characteristic of u. There exists $C>0$ such that, for $y=\gamma(t)$ with $t<0$ and $h \in \mathbb{R}^n$,*

$$|u(y+h)+u(y-h)-2u(y)| \leq \frac{C}{t^2}|h|^2.$$

Theorem 5.12 (Lipschitz graph theorem)**.** *Assume* (5.1). *Let $u \in \mathrm{Lip}\,(\mathbb{T}^n)$ be a solution to* (5.3). *There exists $C>0$ such that, for all $x,y \in \mathcal{M}_0$,*

$$|Du(y)-Du(x)| \leq C|x-y|.$$

Proof. Fix $x, y \in \mathcal{M}_0$. For $z \in \mathbb{T}^n$ to be chosen,

$$\begin{aligned}
|u(y) - u(x) - Du(x) \cdot (y - x)| &\le C|y - x|^2, \\
|u(z) - u(x) - Du(x) \cdot (z - x)| &\le C|z - x|^2, \\
|u(z) - u(y) - Du(y) \cdot (z - y)| &\le C|z - y|^2.
\end{aligned}$$

Combine these three inequalities and use the triangle inequality to yield

$$|(Du(x) - Du(y)) \cdot (z - y)| \le C(|x - y|^2 + |y - z|^2 + |z - x|^2). \tag{5.12}$$

If $Du(x) = Du(y)$, then we are done. Else, choose z as

$$z = y + |x - y| \frac{Du(x) - Du(y)}{|Du(x) - Du(y)|}.$$

Then, we see that

$$|y - z| = |x - y|, \qquad |z - x| \le |x - y| + |y - z| \le 2|x - y|.$$

Plug these into (5.12) to deduce that

$$|Du(y) - Du(x)| \le C|x - y|.$$

□

Corollary 5.13. *Assume* (5.1)*. Let $u \in \mathrm{Lip}\,(\mathbb{T}^n)$ be a solution to* (5.3)*. Then, the projection map $\pi : \widetilde{\mathcal{M}}_0 \to \mathcal{M}_0$ with $\pi(x, Du(x)) = x$ for $x \in \mathcal{M}_0$ is Lipschitz. The inverse map $\pi^{-1} : \mathcal{M}_0 \to \widetilde{\mathcal{M}}_0$ is also Lipschitz.*

5.2.3. Examples of Mather set.

Definition 5.14 (Reversible Lagrangian)**.** The Lagrangian L is said to be reversible if

$$L(x, v) = L(x, -v) \qquad \text{for all } (x, v) \in \mathbb{T}^n \times \mathbb{R}^n.$$

An example of a reversible Lagrangian is

$$L(x, v) = \frac{1}{2}|v|^2 - V(x) \qquad \text{for all } (x, v) \in \mathbb{T}^n \times \mathbb{R}^n,$$

for a given potential energy $V \in C(\mathbb{T}^n)$.

Proposition 5.15. *Assume* (5.1)*. Assume further that L is reversible. Then, the following points hold.*

(i)
$$-c[0] = \min_{x \in \mathbb{T}^n} L(x, 0) = \min_{(x,v) \in \mathbb{T}^n \times \mathbb{R}^n} L(x, v).$$

(ii)
$$\widetilde{\mathcal{M}}_0 = \{(x, 0) \, : \, L(x, 0) = -c[0]\}.$$

Proof. Thanks to (5.1) and the reversibility of L, for $(x,v) \in \mathbb{T}^n \times \mathbb{R}^n \setminus \{0\}$,

$$L(x,v) = \frac{1}{2}\left(L(x,v) + L(x,-v)\right) > L(x,0).$$

Therefore,

$$\min_{x\in\mathbb{T}^n} L(x,0) = \min_{(x,v)\in\mathbb{T}^n\times\mathbb{R}^n} L(x,v) = \alpha \in \mathbb{R}.$$

By Theorem 5.2,

$$\begin{aligned} -c[0] &= \inf\left\{\int_{\mathbb{T}^n\times\mathbb{R}^n} L(x,v)\, d\mu(x,v) \,:\, \mu \in \mathcal{P}(\mathbb{T}^n\times\mathbb{R}^n) \text{ is flow invariant}\right\} \\ &\geq \min_{(x,v)\in\mathbb{T}^n\times\mathbb{R}^n} L(x,v) = \alpha. \end{aligned}$$

On the other hand, for each $x_0 \in \mathbb{T}^n$ such that $L(x_0,0) = \alpha$, we claim that the stay put curve

$$\gamma(t) = x_0 \qquad \text{for all } t \in \mathbb{R}$$

is a minimizing extremal curve. Indeed, for any $\eta \in \mathrm{AC}\left([a,b],\mathbb{T}^n\right)$ with $\eta(a) = \eta(b) = x_0$,

$$\int_a^b L(\gamma,\dot\gamma)\, ds = (b-a)L(x_0,0) \leq \int_a^b L(\eta,\dot\eta)\, ds.$$

Thus,

$$\phi_t^L(x_0,0) = (x_0,0) \qquad \text{for all } t \in \mathbb{R},$$

and hence, $\delta_{(x_0,0)}$ is invariant by ϕ_t^L. As

$$\int_{\mathbb{T}^n\times\mathbb{R}^n} L(x,v)\, d\delta_{(x_0,0)} = \alpha,$$

we conclude that $\delta_{(x_0,0)}$ is a Mather measure. We hence get both (i) and (ii). $\square$

Remark 5.16. Some comments are in order.

(1) The above proposition can be generalized to more complicated cases. For example, we only need to require the Lagrangian satisfying that

$$L(x,v) \geq L(x,0) \qquad \text{for all } (x,v) \in \mathbb{T}^n \times \mathbb{R}^n.$$

(2) In the above proposition, for $x_1,\ldots,x_k \in \arg\min L(\cdot,0)$,

$$\alpha_1\delta_{(x_1,0)} + \cdots + \alpha_k\delta_{(x_k,0)}$$

is a Mather measure for $\alpha_1,\ldots,\alpha_k \geq 0$ and $\sum_{i=1}^k \alpha_i = 1$. In this situation, we say that $\delta_{(x_i,0)}$ is an ergodic Mather measure for $1 \leq i \leq k$.

Let us consider further a more specific example in one dimension.

Example 5.17. Assume $n = 1$, and

$$H(x,p) = \frac{1}{2}|p|^2 + V(x) \qquad \text{for } (x,p) \in \mathbb{T} \times \mathbb{R}.$$

Then,

$$L(x,v) = \frac{1}{2}|v|^2 - V(x) \qquad \text{for } (x,v) \in \mathbb{T} \times \mathbb{R}.$$

As proved in the previous proposition,

$$\begin{cases} c[0] = \max_{\mathbb{T}} V, \\ \widetilde{\mathcal{M}}_0 = \{(x,0) \,:\, V(x) = c[0] = \max_{\mathbb{T}} V\}. \end{cases}$$

Let us now consider the simplest case in which

$$\widetilde{\mathcal{M}}_0 = \{(0,0)\}.$$

We already proved that $\mathcal{M}_0 = \{0\}$ is the uniqueness set for solutions to (5.3). Let us now construct all possible viscosity solutions to (5.3). The PDE for u is

$$\frac{1}{2}|u'(x)|^2 + V(x) = c[0] = \max V \qquad \text{in } \mathbb{T}. \tag{5.13}$$

Then, for a.e. $x \in [0,1]$,

$$u'(x) = \pm\sqrt{2(c[0] - V(x))}.$$

Choose $z \in (0,1)$ such that

$$\int_0^z \sqrt{2(c[0] - V(x))}\, dx = \int_z^1 \sqrt{2(c[0] - V(x))}\, dx.$$

Set

$$u(x) = \begin{cases} \int_0^x \sqrt{2(c[0] - V(s))}\, ds & \text{for } 0 \le x \le z, \\ \int_x^1 \sqrt{2(c[0] - V(s))}\, ds & \text{for } z \le x \le 1. \end{cases}$$

Then,

$$u'(x) = \begin{cases} \sqrt{2(c[0] - V(x))} & \text{for } 0 \le x < z, \\ -\sqrt{2(c[0] - V(x))} & \text{for } z < x \le 1. \end{cases}$$

It is clear that

$$u(0) = u'(0) = u(1) = u'(1) = 0,$$

and u is not differentiable at z. Extend u in a periodic way to $\mathbb{R}$.

We now show quickly that u is a viscosity solution to (5.13). We only need to verify this at z. It is clear that

$$D^+u(z) = [-\sqrt{2(c[0] - V(z))}, -\sqrt{2(c[0] - V(z))}],$$

and for any $p \in D^+u(z)$,

$$\frac{1}{2}|p|^2 + V(z) \le c[0].$$

Thus, all viscosity solutions to (5.13) are $u+C$ for $C\in\mathbb{R}$.

5.3. The Peierls barrier

Definition 5.18 (the Peierls barrier)**.** Define $h:\mathbb{T}^n\times\mathbb{T}^n\to\mathbb{R}$ as: For $x,y\in\mathbb{T}^n$,

$$h(x,y)=\liminf_{t\to\infty}\left[h_t(x,y)+c[0]t\right].$$

Recall that $h_t(x,y)$ is the minimal cost it takes to travel from x to y in a given fixed amount of time t corresponding to the given Lagrangian L. More specifically, as defined in (4.13),

$$h_t(x,y)=\inf_{\substack{\gamma\in\mathrm{AC}\,([0,t],\mathbb{T}^n)\\ \gamma(0)=x,\gamma(t)=y}}\int_0^t L(\gamma(s),\dot\gamma(s))\,ds.$$

Lemma 5.19 (Properties of h_t)**.** *Assume* (5.1)*. Then, the following points hold.*

(1) *For $x,y,z\in\mathbb{T}^n$ and $t,t'>0$,*
$$h_t(x,y)+h_{t'}(y,z)\ge h_{t+t'}(x,z).$$
(2) *If $u\prec L+c$ for some $u\in C(\mathbb{T}^n)$ and $c\in\mathbb{R}$, then, for $x,y\in\mathbb{T}^n$ and $t>0$,*
$$u(y)-u(x)\le h_t(x,y)+ct.$$
(3) *For $x\in\mathbb{T}^n$ and $t>0$, we have*
$$h_t(x,x)+c[0]t\ge 0.$$
(4) *For each $u\in\mathcal{S}_-$ and $t_0>0$, there exists a constant $C=C(u,t_0)>0$ such that, for $x,y\in\mathbb{T}^n$ and $t>t_0$,*
$$-2\|u\|_{L^\infty(\mathbb{T}^n)}\le h_t(x,y)+c[0]t\le 2\|u\|_{L^\infty(\mathbb{T}^n)}+C.$$
(5) *For each $t>0$ and $x,y\in\mathbb{T}^n$, there exists an extremal curve $\gamma:[0,t]\to\mathbb{T}^n$ with $\gamma(0)=x$, $\gamma(t)=y$ such that*
$$h_t(x,y)=\int_0^t L(\gamma(s),\dot\gamma(s))\,ds.$$
Moreover, an extremal curve $\gamma:[0,t]\to\mathbb{T}^n$ is minimizing if and only if
$$h_t(\gamma(0),\gamma(t))=\int_0^t L(\gamma(s),\dot\gamma(s))\,ds.$$
(6) *For each $t_0>0$, there exists a constant $C=C(t_0)>0$ such that, for each $t>t_0$, h_t is Lipschitz in $\mathbb{T}^n\times\mathbb{T}^n$ with Lipschitz constant at most $C(t_0)$.*

Proof. We note that we already proved most of the claims. More precisely, (1), (2), and (5) were shown in the proof of Proposition 4.32. Claim (6) was obtained in Theorem 4.34.

Let us proceed to prove (3). Take $u \in \mathcal{S}_-$. Then, in light of (2),

$$0 = u(x) - u(x) \leq h_t(x, x) + c[0]t.$$

Finally, we prove (4). The lower bound is rather obvious as

$$-2\|u\|_{L^\infty(\mathbb{T}^n)} \leq u(y) - u(x) \leq h_t(x, y) + c[0]t.$$

The upper bound is important as we need to have that for all $t > t_0$. First of all, it is clear that, for each $x, z \in \mathbb{T}^n$, we can find $\gamma^{x,z} : [0, t_0] \to \mathbb{T}^n$ such that $\gamma^{x,z}(0) = x$, $\gamma^{x,z}(t_0) = z$, and

$$h_{t_0}(x, z) = \int_0^{t_0} L(\gamma^{x,z}, \dot{\gamma}^{x,z})\, ds \leq C = C(t_0).$$

Secondly, as $u \in \mathcal{S}_-$, we can find a calibrated curve $\xi : (-\infty, 0] \to \mathbb{T}^n$ such that $\xi(0) = y$, and for $t_2 < t_1 \leq 0$,

$$u(\xi(t_1)) - u(\xi(t_2)) = \int_{t_2}^{t_1} \Big(L(\xi, \dot{\xi}) + c[0] \Big)\, ds.$$

We now combine these two points to conclude. Let $z = \xi(t_0 - t)$. Define $\gamma : [0, t] \to \mathbb{T}^n$ connecting x to y as

$$\gamma(s) = \begin{cases} \gamma^{x,z}(s) & \text{for } 0 \leq s \leq t_0, \\ \xi(s - t) & \text{for } t_0 \leq s \leq t. \end{cases}$$

Then,

$$\begin{aligned} h_t(x, y) + c[0]t &\leq \int_0^t \left(L(\gamma, \dot{\gamma}) + c[0] \right)\, ds \\ &\leq (C(t_0) + c[0]t_0) + u(\xi(0)) - u(\xi(t_0 - t)) \\ &\leq C(t_0) + 2\|u\|_{L^\infty(\mathbb{T}^n)}. \end{aligned}$$

□

Theorem 5.20 (Properties of the Peierls barrier)**.** *Assume* (5.1)*. Then, the following points hold.*

(1) *h is Lipschitz.*

(2) *If $u \prec L + c[0]$ for $u \in C(\mathbb{T}^n)$, then, $x, y \in \mathbb{T}^n$,*

$$u(y) - u(x) \leq h(x, y).$$

(3) *For $x \in \mathbb{T}^n$, $h(x, x) \geq 0$.*

(4) *For $x, y, z \in \mathbb{T}^n$, we have the triangle inequality*

$$h(x, y) + h(y, z) \geq h(x, z).$$

(5) *For* $x, y \in \mathbb{T}^n$,

$$h(x,y) + h(y,x) \geq 0.$$

(6) *For* $x \in \mathcal{M}_0$, $h(x,x) = 0$.

(7) *For* $x, y \in \mathbb{T}^n$, *there exists a sequence of minimizing extremal curves* $\gamma_k : [0, t_k] \to \mathbb{T}^n$ *with* $t_k \to \infty$, $\gamma_k(0) = x$, $\gamma_k(t_k) = y$, *and*

$$h(x,y) = \lim_{k\to\infty} \int_0^{t_k} (L(\gamma_k, \dot{\gamma}_k) + c[0])\, ds.$$

(8) *For any sequence of continuous piecewise* C^1 *curve* $\gamma_k : [0, t_k] \to \mathbb{T}^n$ *with* $t_k \to \infty$, $\gamma_k(0) \to x$, $\gamma_k(t_k) \to y$, *we have*

$$h(x,y) \leq \liminf_{k\to\infty} \int_0^{t_k} (L(\gamma_k, \dot{\gamma}_k) + c[0])\, ds.$$

Proof. Note first that h is finite by item (4) of Lemma 5.19.

For $t \geq 1$, h_t is Lipschitz in $\mathbb{T}^n \times \mathbb{T}^n$ with Lipschitz constant at most $C(1)$. As such, h is Lipschitz with Lipschitz constant at most $C(1)$, and (1) is proved.

Point (2) is rather straightforward as for $x, y \in \mathbb{T}^n$ and $t > 0$,

$$u(y) - u(x) \leq h_t(x,y) + c[0]t.$$

Take lim inf of the above as $t \to \infty$ to conclude. We then take $x = y$ in (2) to get (3).

By item (1) of Lemma 5.19, for $x, y, z \in \mathbb{T}^n$ and $t, t' > 0$,

$$h_t(x,y) + h_{t'}(y,z) \geq h_{t+t'}(x,z).$$

Hence,

$$(h_t(x,y) + c[0]t) + (h_{t'}(y,z) + c[0]t') \geq h_{t+t'}(x,z) + c[0](t+t').$$

Take lim inf of the above left hand side as $t \to \infty$ and $t' \to \infty$ to imply (4). Item (5) follows immediately from (4).

Let us now prove (6), which is very interesting as we start seeing connections between points in the projected Mather set $\mathcal{M}_0$ and the Peierls barrier. Take $x \in \mathcal{M}_0$. There is $v \in \mathbb{R}^n$ such that $(x,v) \in \widetilde{\mathcal{M}}_0$. Pick μ to be a Mather measure such that $(x,v) \in \operatorname{supp}(\mu)$. Note that the recurrent points of ϕ_t^L contained in $\operatorname{supp}(\mu)$ form a dense set in $\operatorname{supp}(\mu)$. By the continuity of h, we can then assume that (x,v) is a recurrent point of ϕ_t^L. In particular, there exists a sequence $\{t_k\} \to \infty$ such that

$$\lim_{k\to\infty} \pi \circ \phi_{t_k}(x,v) = x. \tag{5.14}$$

Fix $u \in \mathcal{S}_-$. We have

$$u(\pi \circ \phi^L_{t_k}(x,v)) - u(x) = \int_0^{t_k} \left(L(\phi^L_s(x,v)) + c[0]\right) ds,$$

which together with (5.14) yields

$$\lim_{k\to\infty} \int_0^{t_k} \left(L(\phi^L_s(x,v)) + c[0]\right) ds = 0. \tag{5.15}$$

We now construct a sequence of loops connecting x to itself. For each $k \in \mathbb{N}$, if $\pi \circ \phi_{t_k}(x,v) = x$, then denote by $\eta_k(s) = \pi \circ \phi^L_s(x,v)$ for $0 \le s \le t_k$. Else, let $s_k = t_k + |\pi \circ \phi_{t_k}(x,v) - x|$ and

$$\eta_k(s) = \begin{cases} \pi \circ \phi^L_s(x,v) & \text{for } 0 \le s \le t_k, \\ \pi \circ \phi_{t_k}(x,v) + (s - t_k)\dfrac{x - \pi \circ \phi_{t_k}(x,v)}{|x - \pi \circ \phi_{t_k}(x,v)|} & \text{for } t_k \le s \le s_k. \end{cases}$$

In light of (5.14) and (5.15), we see that

$$\lim_{k\to\infty} \int_0^{t_k} \left(L(\eta_k, \dot\eta_k) + c[0]\right) ds = 0,$$

which gives $h(x,x) = 0$.

Item (7) follows directly from the definition of h and item (5) of Lemma 5.19.

Finally, let us prove (8). We use a similar construction to the one of η_k above. The point is that, as $k \to \infty$, the costs of connecting x to $\gamma_k(0)$ and $\gamma_k(t_k)$ to y vanish. To be more precise, let $\alpha_k = |x - \gamma_k(0)|$, and $\beta_k = |\gamma_k(t_k) - y|$. Set

$$\xi_k(s) = \begin{cases} x + s\frac{\gamma_k(0) - x}{|\gamma_k(0) - x|} & \text{for } 0 \le s \le \alpha_k, \\ \gamma_k(s - \alpha_k) & \text{for } \alpha_k \le s \le t_k + \alpha_k, \\ \gamma_k(t_k) + (s - (t_k + \alpha_k))\frac{y - \gamma_k(t_k)}{|y - \gamma_k(t_k)|} & \text{for } t_k + \alpha_k \le s \le t_k + \alpha_k + \beta_k. \end{cases}$$

Then,

$$\begin{aligned} h(x,y) &\le \liminf_{k\to\infty} \int_0^{t_k+\alpha_k+\beta_k} \left(L(\xi_k, \dot\xi_k) + c[0]\right) ds \\ &\le \liminf_{k\to\infty} \int_0^{t_k} \left(L(\gamma_k, \dot\gamma_k) + c[0]\right) ds. \end{aligned}$$

The proof is complete. $\square$

We next have the following important lemma.

Lemma 5.21. *Assume* (5.1). *Let V be an open neighborhood of $\widetilde{\mathcal{M}}_0$ in $\mathbb{T}^n \times \mathbb{R}^n$. Then, there exists $T = T(V) > 0$ such that if $\gamma : [0,t] \to \mathbb{T}^n$ is a minimizing curve with $t \ge T$, then we can find $s \in [0,t]$ such that $(\gamma(s), \dot\gamma(s)) \in V$.*

Proof. We give a proof by contradiction. Assume otherwise that we can find $\{t_k\} \to \infty$ and a sequence of minimizing curves $\gamma_k : [0, t_k] \to \mathbb{T}^n$ such that

$$\{(\gamma_k(s), \dot{\gamma}_k(s)) \,:\, 0 \le s \le t_k\} \cap \overline{V} = \emptyset. \tag{5.16}$$

Without loss of generality, assume $t_k \ge 1$ for all $k \in \mathbb{N}$ and there exists a compact set $K \subset \mathbb{R}^n$ such that, for all $k \in \mathbb{N}$,

$$\{(\gamma_k(s), \dot{\gamma}_k(s)) \,:\, 0 \le s \le t_k\} \subset \mathbb{T}^n \times K. \tag{5.17}$$

We now construct a Mather measure from $\{\gamma_k\}$ to get a contradiction. Let $\mu_k \in \mathcal{P}(\mathbb{T}^n \times \mathbb{R}^n)$ be such that

$$\int_{\mathbb{T}^n \times \mathbb{R}^n} \psi(x,v)\, d\mu_k(x,v) = \frac{1}{t_k} \int_0^{t_k} \psi(\gamma_k(s), \dot{\gamma}_k(s))\, ds$$

for all ψ continuous and bounded in $\mathbb{T}^n \times \mathbb{R}^n$. By (5.17), $\mathrm{supp}\,(\mu_k) \subset \mathbb{T}^n \times K$. By passing to a subsequence if necessary, we assume that $\mu_k \rightharpoonup \mu$ weakly in the sense of measure for some $\mu \in \mathcal{P}(\mathbb{T}^n \times \mathbb{R}^n)$. It is clear that $\mathrm{supp}\,(\mu) \subset \mathbb{T}^n \times K$ and μ is invariant under ϕ_t^L. Besides, for each $k \in \mathbb{N}$,

$$\int_{\mathbb{T}^n \times \mathbb{R}^n} L(x,v)\, d\mu_k(x,v) = \frac{1}{t_k} h_{t_k}(\gamma_k(0), \gamma_k(t_k)),$$

which together with item (4) of Lemma 5.19 implies

$$-\frac{2\|u\|_{L^\infty(\mathbb{T}^n)}}{t_k} \le \int_{\mathbb{T}^n \times \mathbb{R}^n} L(x,v)\, d\mu_k(x,v) + c[0] \le \frac{2\|u\|_{L^\infty(\mathbb{T}^n)} + C}{t_k}.$$

Let $k \to \infty$ to deduce that

$$\int_{\mathbb{T}^n \times \mathbb{R}^n} L(x,v)\, d\mu(x,v) = -c[0],$$

which gives further that μ is a Mather measure. This contradicts (5.16).

□

5.4. Aubry set

There are many different ways to define Aubry set. We give here one that uses h.

Definition 5.22 (Aubry set)**.** Assume (5.1). Denote the Aubry set $\mathcal{A}_0$ as

$$\mathcal{A}_0 = \{x \in \mathbb{T}^n \,:\, h(x,x) = 0\}.$$

It is clear that $\mathcal{A}_0 \neq \emptyset$ as item (6) of Theorem 5.20 gives

$$\emptyset \neq \mathcal{M}_0 \subset \mathcal{A}_0.$$

We have the following clear characterizations of $\mathcal{A}_0$.

Proposition 5.23. *Assume* (5.1)*. The followings are equivalent.*

(i) $x \in \mathcal{A}_0$, *that is,* $h(x,x) = 0$.

(ii) *There exists a sequence* $\{\gamma_k\}$ *of continuous piecewise* C^1 *curves* $\gamma_k : [0,t_k] \to \mathbb{T}^n$ *with* $\gamma_k(0) = \gamma_k(t_k) = x$ *and* $t_k \to \infty$ *such that*

$$\lim_{k\to\infty} \int_0^{t_k} \left(L(\gamma_k, \dot{\gamma}_k) + c[0]\right)\, ds = 0.$$

(iii) *There exists a sequence* $\{\gamma_k\}$ *of minimizing extremal curves* γ_k : $[0,t_k] \to \mathbb{T}^n$ *with* $\gamma_k(0) = \gamma_k(t_k) = x$ *and* $t_k \to \infty$ *such that*

$$\lim_{k\to\infty} \int_0^{t_k} \left(L(\gamma_k, \dot{\gamma}_k) + c[0]\right)\, ds = 0.$$

The proof of this proposition is straightforward and hence is omitted. Next, we give another characterization of the Aubry set.

Theorem 5.24. *Assume* (5.1). *Then,* $x \in \mathcal{A}_0$ *if and only if for any fixed* $\delta > 0$,

$$\inf\left\{ \int_0^t \left(L(\gamma, \dot{\gamma}) + c[0]\right)\, ds \ : \ \gamma \in \mathrm{AC}\left([0,t], \mathbb{T}^n\right) \text{ with } t > \delta, \gamma(0) = \gamma(t) = x \right\} = 0.$$

Proof. We first prove the "⇒" direction. Take $x \in \mathcal{A}_0$. By (ii) of Proposition 5.23, there exists a sequence $\{\gamma_k\}$ of continuous piecewise C^1 curves $\gamma_k : [0,t_k] \to \mathbb{T}^n$ with $\gamma_k(0) = \gamma_k(t_k) = x$ and $t_k \to \infty$ such that

$$\lim_{k\to\infty} \int_0^{t_k} \left(L(\gamma_k, \dot{\gamma}_k) + c[0]\right)\, ds = 0,$$

which allows us to conclude right away.

Next, we prove the "⇐" direction. Fix $\delta > 0$. For each $k \in \mathbb{N}$, there exists $\gamma \in \mathrm{AC}\left([0,t], \mathbb{T}^n\right)$ with $t > \delta$ and $\gamma(0) = \gamma(t) = x$ such that

$$0 \le \int_0^t \left(L(\gamma, \dot{\gamma}) + c[0]\right)\, ds \le \frac{1}{k^2}.$$

Let γ_k be k copies of γ (or γ with multiplicity k). More explicitly, let $t_k = kt$, and

$$\gamma_k(s) = \gamma(s - it) \qquad \text{for } i \in \{0, \ldots, k-1\}, it \le s \le (t+1)t.$$

Then,

$$0 \le \int_0^{t_k} \left(L(\gamma_k, \dot{\gamma}_k) + c[0]\right)\, ds = 0 \le k\frac{1}{k^2} = \frac{1}{k}.$$

Besides, $t_k = kt \ge k\delta$, and so, $\lim_{k\to\infty} t_k = \infty$. By (ii) of Proposition 5.23, we yield that $x \in \mathcal{A}_0$. $\square$

Remark 5.25. Note that for $x \in \mathcal{A}_0$, we do not know in general if there exists a loop $\gamma : [0,t] \to \mathbb{T}^n$ with $t > 0$, $\gamma(0) = \gamma(t) = x$ such that

$$\int_0^t \left(L(\gamma, \dot\gamma) + c[0]\right)\, ds = 0.$$

Example 5.26. Consider the case where

$$H(x,p) = \frac{1}{2}|p|^2 + V(x) \qquad \text{for } (x,p) \in \mathbb{T}^n \times \mathbb{R}^n,$$

for some $V \in C(\mathbb{T}^n)$. Then,

$$L(x,v) = \frac{1}{2}|v|^2 - V(x) \qquad \text{for } (x,v) \in \mathbb{T}^n \times \mathbb{R}^n.$$

We already computed that $c[0] = \max V$, and

$$\mathcal{M}_0 = \{y \in \mathbb{T}^n \,:\, V(y) = c[0] = \max V\}.$$

Let us find out what is $\mathcal{A}_0$ in this case. Note that for $\gamma : [0,t] \to \mathbb{T}^n$ with $t > 0$, and $\gamma(0) = \gamma(t) = x$,

$$\begin{aligned} &\int_0^t \left(L(\gamma, \dot\gamma) + c[0]\right)\, ds \\ &= \int_0^t \left(\frac{1}{2}|\dot\gamma(s)|^2 + \max V - V(\gamma(s))\right) ds \geq 0. \end{aligned}$$

It is clear that equality in the above happens if and only if

$$\gamma(s) = x \in \mathcal{M}_0 \qquad \text{for all } s \in [0,t].$$

Next, take $x \notin \mathcal{M}_0$. Then, $V(x) = c[0] - 2\theta < c[0]$ for some $\theta > 0$. Let

$$O = V^{-1}([c[0] - \theta, c[0]]) \supset \mathcal{M}_0.$$

Then, $\delta = \mathrm{dist}\,(x, O) > 0$. For any $t > 1$, denote by $J = \{s \in [0,t] \,:\, \gamma(s) \in O\}$, and

$$\tau = \begin{cases} \inf J & \text{if } J \neq \emptyset, \\ t & \text{if } J = \emptyset. \end{cases}$$

Then,

$$\begin{aligned} &\int_0^t \left(L(\gamma, \dot\gamma) + c[0]\right)\, ds \\ &\geq \int_0^\tau \left(\frac{1}{2}|\dot\gamma(s)|^2 + \theta\right) ds \\ &\geq \theta\tau + \frac{\tau}{2}\left|\frac{1}{\tau}\int_0^\tau \dot\gamma(s)\, ds\right|^2 = \theta\tau + \frac{|\gamma(\tau) - x|^2}{2} \\ &\geq \min\left\{\theta, \delta\sqrt{2\theta}\right\}, \end{aligned}$$

which implies that $x \notin \mathcal{A}_0$. We hence conclude in this case that

$$\mathcal{A}_0 = \mathcal{M}_0 = \{y \in \mathbb{T}^n \,:\, V(y) = c[0] = \max V\}.$$

We discuss further about properties of h.

Theorem 5.27. *Assume* (5.1). *Fix* $x \in \mathbb{T}^n$. *Denote by*

$$h^x(y) = h(x,y) \qquad \text{for } y \in \mathbb{T}^n.$$

Then, $h^x \in \mathcal{S}_-$.

Proof. We first show that $h^x \prec L + c[0]$. Fix $\gamma \in \mathrm{AC}\,([0,t],\mathbb{T}^n)$. Of course,

$$h_t(\gamma(0),\gamma(t)) \le \int_0^t L(\gamma,\dot\gamma)\,ds,$$

and, for $t' > 0$, by the triangle inequality,

$$h_{t+t'}(x,\gamma(t)) \le h_{t'}(x,\gamma(0)) + \int_0^t L(\gamma,\dot\gamma)\,ds.$$

Therefore,

$$h_{t+t'}(x,\gamma(t)) + c[0](t+t') \le h_{t'}(x,\gamma(0)) + c[0]t' + \int_0^t L(\gamma,\dot\gamma)\,ds + c[0]t.$$

Take $\liminf$ as $t' \to \infty$ to yield

$$h^x(\gamma(t)) - h^x(\gamma(0)) \le \int_0^t L(\gamma,\dot\gamma)\,ds + c[0]t. \tag{5.18}$$

We get $h^x \prec L + c[0]$.

Next, to finish the proof, we need to show that for each $y \in \mathbb{T}^n$, there exists a calibrated curve $\gamma : (-\infty,0] \to \mathbb{T}^n$ with $\gamma(0) = y$, and for $t > 0$,

$$h^x(\gamma(0)) - h^x(\gamma(-t)) = \int_{-t}^0 L(\gamma,\dot\gamma)\,ds + c[0]t.$$

Take a sequence of extremal curves $\eta_k : [-t_k,0] \to \mathbb{T}^n$ connecting x to y with $t_k \to \infty$, $\eta_k(-t_k) = x$, $\eta_k(0) = y$, and

$$h^x(y) = h(x,y) = \lim_{k\to\infty}\left(\int_{-t_k}^0 L(\eta_k,\dot\eta_k)\,ds + c[0]t_k\right)$$

There exists $C > 0$ such that, for all $k \in \mathbb{N}$,

$$\|\eta_k\|_{L^\infty([-t_k,0])} + \|\dot\eta_k\|_{L^\infty([-t_k,0])} \le C.$$

By a diagonal argument and the Arzelà-Ascoli theorem, by passing to a subsequence if needed, there exists $\gamma \in \mathrm{Lip}\,((-\infty,0],\mathbb{T}^n)$ with $\gamma(0) = y$ such that

$$\eta_k \to \gamma \qquad \text{locally uniformly on } (-\infty,0].$$

We claim that γ is exactly the calibrated curve that we want. Indeed, fix $t>0$. For k large enough such that $t_k > t$, then

$$\begin{aligned}&\int_{-t_k}^{0} L(\eta_k,\dot\eta_k)\,ds + c[0]t_k\\ &= \int_{-t_k}^{-t} L(\eta_k,\dot\eta_k)\,ds + c[0](t_k-t) + \int_{-t}^{0} L(\eta_k,\dot\eta_k)\,ds + c[0]t.\end{aligned}$$

Let $k\to\infty$ to imply

$$h^x(y) \geq h^x(\gamma(-t)) + \int_{-t}^{0} L(\gamma,\dot\gamma)\,ds + c[0]t. \tag{5.19}$$

Combine (5.18) and (5.19) to conclude the proof. $\square$

5.5. References

(1) The cell problems were first studied by Lions, Papanicolaou, Varadhan [**LPV**].

(2) For the weak KAM theorem via dynamical viewpoint, see Fathi [**Fat**]. In fact, this chapter is heavily based on [**Fat**]. See also the books of Gomes [**Gom09**], Sorrentino [**Sor15**], Tran [**Tra21**].

(3) There are many excellent survey papers and lecture notes in weak KAM theory: see Evans [**Eva08, Eva04**], Ishii [**Ish**], Kaloshin [**Kal05**], and the references therein.

Chapter 6

Aubry-Mather theory in two dimensions in the smooth setting

In this chapter, we are always in two dimensions. We consider a given Hamiltonian $H : \mathbb{T}^2 \times \mathbb{R}^2 \to \mathbb{R}$ that satisfies

$$\begin{cases} H \in C^k(\mathbb{T}^2 \times \mathbb{R}^2) \text{ for some } k \geq 2, \\ D^2_{pp}H(y,p) > 0 \text{ for all } (y,p) \in \mathbb{T}^2 \times \mathbb{R}^2, \\ \lim_{|p|\to\infty} \min_{y\in\mathbb{T}^2} \frac{H(y,p)}{|p|} = +\infty. \end{cases} \tag{6.1}$$

Let L be the corresponding Lagrangian (the Legendre transform of H). Then, L satisfies

$$\begin{cases} L \in C^k(\mathbb{T}^2 \times \mathbb{R}^2), \\ D^2_{vv}L(y,v) > 0 \text{ for all } (y,v) \in \mathbb{T}^2 \times \mathbb{R}^2, \\ \lim_{|v|\to\infty} \min_{y\in\mathbb{T}^2} \frac{L(y,v)}{|v|} = +\infty. \end{cases} \tag{6.2}$$

The main object in this chapter is still the cell problem at $p \in \mathbb{R}^2$, that is,

$$H(y, p + Dv(y)) = \overline{H}(p) \qquad \text{in } \mathbb{T}^2. \tag{6.3}$$

Here, $\overline{H}(p) \in \mathbb{R}$ is the unique constant so that (6.3) has a viscosity solution as discussed in the previous chapters.

6.1. Absolute minimizing curves

6.1.1. Absolute minimizing curves.

Definition 6.1 (Absolute minimizing curve)**.** Assume (6.1). A curve $\gamma \in \mathrm{AC}(\mathbb{R}, \mathbb{R}^2)$ is called an absolute minimizer (or absolute minimizing curve) associated with $L + c$ for some $c \in \mathbb{R}$ if for any $t_1 < t_2$,

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

for every $\delta \in \mathrm{AC}([s_1, s_2], \mathbb{R}^2)$ 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 [**Ban88**]).

Theorem 6.2. *Assume* (6.1)*. Let γ_1 and γ_2 be two distinct (up to translation in time) absolute minimizers associated with $L + c$ for some given $c \in \mathbb{R}$. Then, γ_1 and γ_2 intersect at most once.*

Proof. Assume otherwise that γ_1 and γ_2 intersect at least twice. By a suitable translation in time, we may assume that there are $a, b_1, b_2 \in \mathbb{R}$ such that $a < b_1 \le b_2$ and

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

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

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

As

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

we yield that γ_3 is also a minimizer of the action

$$\int_a^{b_1+1} (L(\gamma, \dot\gamma) + c)\ ds$$

with corresponding fixed endpoints $\gamma(a) = \gamma_2(a)$ and $\gamma(b_1 + 1) = \gamma_2(b_2 + 1)$. Hence γ_3 is C^k and solves the following Euler-Lagrange equations

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

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

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

Since γ_1 and γ_2 are also solutions to the above Euler-Lagrange equations, the uniqueness result for second order ODEs yields that $\gamma_1(t) = \gamma_2(t + b_2 - b_1)$, which is absurd.

□

6.1.2. Preliminaries on orbits in projected Mather sets. Let us recall the result in Lemma 5.5. Let $u \in \mathrm{Lip}(\mathbb{T}^2)$ be a solution to (6.3). Pick $(x,v) \in \widetilde{\mathcal{M}}_p$. Then, for $t \le t'$,

$$\begin{aligned} p \cdot (\pi \circ \phi^L_{t'}(x,v)) + u\left(\pi \circ \phi^L_{t'}(x,v)\right) - p \cdot (\pi \circ \phi^L_t(x,v)) \\ - u\left(\pi \circ \phi^L_t(x,v)\right) = \int_t^{t'} \left(L\left(\phi^L_s(x,v)\right) + \overline{H}(p)\right) ds. \end{aligned} \tag{6.4}$$

Definition 6.3 (Orbits in projected Mather sets)**.** Assume (6.1). Fix $p \in \mathbb{R}^2$. For each $(x,v) \in \widetilde{\mathcal{M}}_p$, we lift $\pi \circ \phi^L_t(x,v)$ for $t \in \mathbb{R}$ to $\mathbb{R}^2$ and say that it is an orbit in $\mathcal{M}_p$.

By using a similar idea to that in the proof of Theorem 6.2, we also have the following result.

Theorem 6.4. *Assume* (6.1)*. Let γ_1 and γ_2 be two distinct (up to translation in time) orbits in $\mathcal{M}_p$ for some $p \in \mathbb{R}^2$. Then, γ_1 and γ_2 do not intersect.*

Proof. Without loss of generality, assume $p = 0$. Assume otherwise that γ_1 and γ_2 intersect at least once. By a suitable translation in time, we may assume that there exists $a \in \mathbb{R}$ such that

$$\gamma_1(a) = \gamma_2(a).$$

Let $\gamma_3 : \mathbb{R} \to \mathbb{R}^2$ be such that

$$\gamma_3(s) = \begin{cases} \gamma_1(s) & \text{for } s \le a, \\ \gamma_2(s) & \text{for } s \ge a. \end{cases}$$

Then, in light of (6.4), γ_3 is an absolute minimizer associated with $L + \overline{H}(0)$. Hence γ_3 is C^k and solves the following Euler-Lagrange equations

$$\frac{d}{ds}\left(D_v L(\gamma_3(s), \dot\gamma_3(s))\right) = D_x L(\gamma_3(s), \dot\gamma_3(s)) \quad \text{for all } s \in \mathbb{R}.$$

Accordingly, at the junction, we have

$$\dot\gamma_1(a) = \dot\gamma_2(a) = \dot\gamma_3(a).$$

Since γ_1 and γ_2 are also solutions to the above Euler-Lagrange equations, the uniqueness result for second order ODEs yields that $\gamma_1 = \gamma_2$, which gives a contradiction.

□

6.1.3. Existence of periodic orbits.

Proposition 6.5. *Assume* (6.1), *and*

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

If $c > 0$, then there exists a periodic orbit $\eta : \mathbb{R} \to \mathbb{R}^2$, which is an absolute minimizer associated with $L + c$, and

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

for some $T > 0$.

Proof. Denote by

$$\Gamma(1) = \{\gamma \in \mathrm{AC}(\mathbb{R}, \mathbb{R}^2) \,:\, \text{there exists } T_\gamma > 0 \text{ such that } \gamma(t + T_\gamma) = \gamma(t) + (0,1) \text{ for } t \in \mathbb{R}\}.$$

We study the following minimization problem

$$\inf_{\gamma \in \Gamma(1)} \int_0^{T_\gamma} (L(\gamma, \dot\gamma) + c)\, ds. \tag{6.5}$$

Let w be a solution to (6.3) with $p = 0$. As $\overline{H}(0) = 0$, for any $\gamma \in \Gamma(1)$,

$$\int_0^{T_\gamma} L(\gamma, \dot\gamma)\, ds \geq w(\gamma(T_\gamma)) - w(\gamma(0)) = 0,$$

and hence,

$$\int_0^{T_\gamma} (L(\gamma, \dot\gamma) + c)\, ds \geq cT_\gamma > 0.$$

Let $\xi(s) = (0, s)$ for $s \in \mathbb{R}$. Then, $T_\xi = 1$, and

$$\int_0^{T_\xi} \left(L(\xi, \dot\xi) + c\right) ds \leq c + \max_{\mathbb{T}^2 \times \overline{B}(0,1)} L \leq C.$$

Therefore, we only need to study (6.5) for $\gamma \in \Gamma(1)$ with $T_\gamma \leq C/c$. Thus, (6.5) admits a minimizer $\eta \in C^k([0, T_\eta])$. By shifting the time, we also have that $\eta \in C^k([s, s + T_\eta])$ for any $s \in \mathbb{R}$, which gives us that $\eta \in C^k(\mathbb{R})$.

Let us now prove that η is an absolute minimizer associated with $L + c$. By using the proof of Theorem 6.2, for any two minimizers of (6.5), they intersect at most once in their periods. Therefore, all the minimizers of (6.5) are well-ordered from left to right on the plane (they can still tangentially touch each other).

For $m \in \mathbb{N}$ with $m \geq 2$, set

$$\Gamma(m) = \{\gamma \in \mathrm{AC}(\mathbb{R}, \mathbb{R}^2) \,:\, \text{there exists } T_\gamma^m > 0 \text{ such that } \gamma(t + T_\gamma^m) = \gamma(t) + (0,m) \text{ for } t \in \mathbb{R}\}.$$

We then consider

$$\inf_{\gamma\in\Gamma(m)} \int_0^{T_\gamma^m} (L(\gamma,\dot\gamma)+c)\, ds. \tag{6.6}$$

By a similar argument to the above, (6.6) admits a minimizer $\eta^m \in C^k(\mathbb{R})$. Again, all the minimizers of (6.6) are well-ordered from left to right on the plane.

We claim that in fact, η^m is also a minimizer to (6.5). Indeed, both η^m and $\eta^m+(0,k)$ for $k\in\mathbb{N}$ are minimizers of (6.6). Without loss of generality, assume η^m is on the left of $\eta^m+(0,1)$. By shifting, $\eta^m+(0,k)$ is on the left of $\eta^m+(0,k+1)$ for $k\in\mathbb{N}$. Since $\eta^m=\eta^m+(0,m)$, we see that $\eta^m=\eta^m+(0,1)$. Thus, our claim holds true. We get further that η is a minimizer to (6.6) for all $m\in\mathbb{N}$. This important point implies that η is an absolute minimizer associated with $L+c$.

□

6.2. Regularity of the level curves of the effective Hamiltonian

Here is the main result of this section.

Theorem 6.6. *Assume* (6.1). *If $\overline{H}(p) > \min_{\mathbb{R}^2}\overline{H}$, then the subgradient set $\partial\overline{H}(p)$ is a radical segment, that is,*

$$\partial\overline{H}(p)=\{rn_p \,:\, r\in[r_1(p),r_2(p)]\}$$

for some unit vector $n_p\in\mathbb{R}^2$ 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 $n\geq 3$.

Proof. Suppose that $q_1,q_2\in\partial\overline{H}(p)$. Then, there exist two Mather measures μ_1 and μ_2 associated with p such that, for $i=1,2$,

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

Accordingly, if q_1 and q_2 are not parallel, then there are two different orbits γ_1 and γ_2 from supports of μ_1 and μ_2 respectively, which intersect each other. We use strongly the two dimensional geometry in this point. In higher dimensions, it is not necessarily the case that γ_1 and γ_2 intersect. However, this is impossible thanks to Theorem 6.4.

Thus, all the vectors in the subgradient set $\partial\overline{H}(p)$ are parallel. In particular, the subgradient set $\partial\overline{H}(p)$ is a radical segment, that is,

$$\partial\overline{H}(p)=\{rn_p \,:\, r\in[r_1(p),r_2(p)]\}$$

for some unit vector $n_p \in \mathbb{R}^2$ and $r_1(p), r_2(p) > 0$. We have furthermore that there exists $s_p \in [-\infty, \infty]$ such that, for any orbit $\xi(t) = (x(t), y(t))$ in the projected Mather set $\mathcal{M}_p$,

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

Fix $s > \min_{\mathbb{R}^2} \overline{H}$. For every $p \in \{\overline{H} = s\}$, the above claim also yields that there is a unique normal vector from p to $\{\overline{H} = s\}$. Therefore, $\{\overline{H} = s\}$ is C^1. $\square$

6.3. Orbits in projected Mather sets

6.3.1. Identification with circle homeomorphisms. Let us first recall a result on circle homeomorphisms.

Definition 6.7 (Circle homeomorphism)**.** A continuous function $f : \mathbb{R} \to \mathbb{R}$ is called a circle homeomorphism if f is strictly increasing and

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

If f is a circle homeomorphism, then it is well-known that the Poincaré rotation number

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

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

$$|f^i(x) - f(x) - i\beta_f| \leq 1 \qquad \text{for all } i \in \mathbb{Z}. \tag{6.7}$$

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.

Now, we identify orbits in projected Mather sets with circle homeomorphisms. Fix $p \in \mathbb{R}^2$. Assume that $\xi : \mathbb{R} \to \mathbb{R}^2$ is an orbit in $\mathcal{M}_p$. By Theorem 6.4, orbits in $\mathcal{M}_p$ do not intersect with each other. Therefore, they are totally ordered in $\mathbb{R}^2$ (see Figure 1 below).

In the following, we explain how to associate $\xi(t) = (x(t), y(t))$ with a circle map $f : \mathbb{R} \to \mathbb{R}$ when $\overline{H}(p) > \min_{\mathbb{R}^2} \overline{H}$, which is well-known in the Aubry-Mather theory. Without loss of generality, we assume that

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

By Proposition 6.5, there exists a periodic trajectory $\eta : \mathbb{R} \to \mathbb{R}^2$, which is an absolute minimizer associated with $L + \overline{H}(p)$, and

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

for some $T > 0$. Clearly, for each $k \in \mathbb{Z}$, ξ intersects with $\eta_k = \eta + (k, 0)$ exactly once since both ξ and η are absolute minimizers of the action

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

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 in $\mathcal{M}_p$ 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 $k > l$,

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

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

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

See [**Ban88**, Theorem 3.15] and Appendix B for further detail on the definition of f.

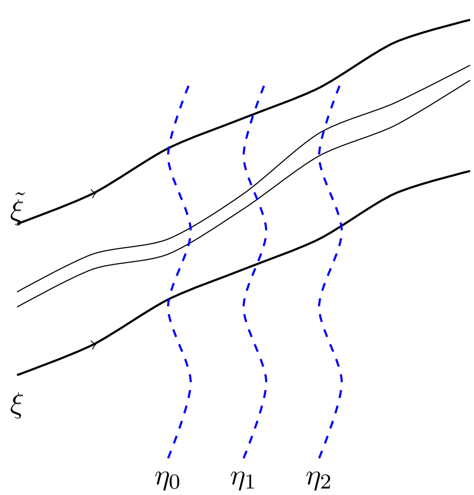

Figure 1. Orbits in $\mathcal{M}_p$ and $\{\eta_k\}_{k\in\mathbb{Z}}$

6.3.2. 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, ξ, and η be from the previous section. Denote

$$\overline{H}(p) = r > 0.$$

By (6.1), there exists $C_r > 0$ such that

$$L(x,v) \geq 2\left(r + \max_{\mathbb{T}^2 \times \overline{B}(0,1)} L\right)|v| - C_r \qquad \text{for all } (x,v) \in \mathbb{T}^2 \times \mathbb{R}^2.$$

We first give uniform lower and upper bounds of the period T of η.

Lemma 6.8. *Assume* (6.1). *Let $\eta : \mathbb{R} \to \mathbb{R}^2$ be a periodic trajectory as in the construction of Proposition 6.5 with $c = r = \overline{H}(p)$, and*

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

for some $T > 0$. Then,

$$\frac{r + \max_{\mathbb{T}^2 \times \overline{B}(0,1)} L}{C_r} \leq T \leq 1 + \frac{\max_{\mathbb{T}^2 \times \overline{B}(0,1)} L}{r}. \tag{6.8}$$

Proof. By the proof of Proposition 6.5,

$$T = T_\eta \leq \frac{r + \max_{\mathbb{T}^2 \times \overline{B}(0,1)} L}{r} = 1 + \frac{\max_{\mathbb{T}^2 \times \overline{B}(0,1)} L}{r}.$$

On the other hand,

$$\begin{aligned} r + \max_{\mathbb{T}^2 \times \overline{B}(0,1)} L &\geq \int_0^T L(\eta, \dot{\eta})\, ds + rT \\ &\geq \int_0^T \left(2\left(r + \max_{\mathbb{T}^2 \times \overline{B}(0,1)} L\right)|\dot{\eta}| - C_r\right) ds \\ &\geq 2\left(r + \max_{\mathbb{T}^2 \times \overline{B}(0,1)} L\right) - C_r T. \end{aligned}$$

Thus,

$$T \geq \frac{r + \max_{\mathbb{T}^2 \times \overline{B}(0,1)} L}{C_r}.$$

□

We give some preparations before stating the main result of this section. Let v be a viscosity solution to

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

Note that v is differentiable along ξ, and

$$\dot{\xi} = D_p H(\xi, p + Dv(\xi)) \qquad \text{in } \mathbb{R}.$$

Therefore, there exists $C(r) > 0$ depending only on r and the growth rate of L such that

$$\|\dot{\xi}\|_{L^\infty(\mathbb{R})},\ \|\dot{\eta}\|_{L^\infty(\mathbb{R})} \leq C(r). \tag{6.9}$$

Let w be a solution to (6.3) with $p = 0$. As $\overline{H}(0) = 0$, for any $\gamma \in \mathrm{AC}\,([s_1, s_2], \mathbb{R}^2)$,

$$\int_{s_1}^{s_2} L(\gamma, \dot{\gamma})\, ds \geq w(\gamma(s_2)) - w(\gamma(s_1)) \geq -d\sqrt{2}, \tag{6.10}$$

where

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

Here is the main result in this section.

Theorem 6.9. *Assume* (6.1). *Assume that* $\xi(t) = (x(t), y(t))$ *for* $t \in \mathbb{R}$ *is an orbit in* $\mathcal{M}_p$, *and*

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

Then for all $t \in \mathbb{R}$,

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

Here, $C > 0$ *is a constant depending only on* r, d, *and the growth rate of* L.

Proof. Assume that

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

In light of (6.8) and (6.9),

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

Therefore,

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

Hence,

$$\lim_{k\to\infty} \frac{a_k}{k} = s_p.$$

Since $t_0 = a_0 = 0$, thanks to (6.7),

$$|a_k - k s_p| \leq 1.$$

Accordingly,

$$|y(t_k) - s_p x(t_k)| \leq C \qquad \text{for all } k \in \mathbb{Z}. \tag{6.11}$$

To finish the proof, we show that

$$|t_k - t_{k+1}| \leq C \qquad \text{for all } k \in \mathbb{Z}. \tag{6.12}$$

In fact, from the previous calculations, we have that

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

which imply

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

We used (6.10) in the last inequality above. Thus, (6.12) holds. Combine (6.9), (6.11) and (6.12) to imply

$$|y(t) - s_p x(t)| \le C \qquad \text{for all } t \in \mathbb{R}. \tag{6.13}$$

□

6.4. Effective fronts in two dimensions

In this section, we consider the Hamiltonian from the front propagation framework

$$\begin{cases} H(y,p) = a(y)|p| \text{ for all } (y,p) \in \mathbb{T}^2 \times \mathbb{R}^2, \\ a \in C^2(\mathbb{T}^2, (0,\infty)). \end{cases} \tag{6.14}$$

Here is the main result of this section.

Theorem 6.10. *Assume* (6.14). *If the level curve* $\left\{\overline{H} = 1\right\}$ *is strictly convex, that is, it does not contain any flat piece, then* a *is constant.*

We list first some basic results in order to prove the above theorem.

6.4.1. Preliminaries.

Lemma 6.11. *Assume* (6.14). *Then,* $\overline{H} \in C^1(\mathbb{R}^2 \setminus \{0\})$.

Proof. Note first that $\overline{H}$ is 1-homogeneous, and

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

Thus, to prove that $\overline{H} \in C^1(\mathbb{R}^2 \setminus \{0\})$, it is enough to show that $\left\{\overline{H} = 1\right\}$ is C^1. The cell problem at p can be written in an equivalent form as

$$a(y)^2 |p + Dv|^2 = \overline{H}(p)^2 \qquad \text{in } \mathbb{T}^2.$$

We then use Theorem 6.6 to conclude. □

Definition 6.12 (Riemannian distance)**.** For $x, y \in \mathbb{R}^2$, denote by $d(x,y)$ the Riemannian distance between x and y, where

$$d(x,y) = \min\left\{ \int_0^1 \frac{|\dot{\xi}(s)|}{a(\xi(s))}\, ds \,:\, \xi \in \mathrm{AC}\,([0,1],\mathbb{R}^2), \xi(0) = x, \xi(1) = y \right\}.$$

In fact, $x \mapsto d^y(x) = d(x,y)$ is the maximal viscosity solution to

$$\begin{cases} a(x)|Dd^y(x)| = 1 & \text{in } \mathbb{R}^2 \setminus \{y\}, \\ d^y(y) = 0. \end{cases}$$

Geometrically, $x \mapsto d^y(x)$ looks like a cone with vertex y.

Definition 6.13 (minimizing geodesics)**.** A curve $c : \mathbb{R} \to \mathbb{R}^2$ is called a minimizing geodesic if, for any $t_1 < t_2$,

$$d(c(t_1), c(t_2)) = \int_{t_1}^{t_2} \frac{|\dot{c}(s)|}{a(c(s))}\, ds.$$

Moreover, $\gamma : [0,\infty) \to \mathbb{R}^2$ is called a minimizing ray (or simply a ray) if the above equality holds for any $0 \le t_1 < t_2$.

Lemma 6.14. *Assume* (6.14)*. For $p_0 \in \left\{\overline{H} = 1\right\}$, let v_0 be a viscosity solution to* (6.3) *with $p = p_0$. Then, any global characteristic of v_0 is a minimizing geodesic with rotation vector $D\overline{H}(p_0)$.*

Proof. Assume that $\gamma : \mathbb{R} \to \mathbb{R}^2$ is a global characteristic of v_0. Then, for $s \in \mathbb{R}$, v_0 is differentiable at $\gamma(s)$, and

$$\dot{\gamma}(s) = a(\gamma(s)) \frac{p_0 + Dv_0(\gamma(s))}{|p_0 + Dv_0(\gamma(s))|}. \tag{6.15}$$

Moreover, for $t_1 < t_2$,

$$p_0 \cdot \gamma(t_2) + v_0(\gamma(t_2)) - p_0 \cdot \gamma(t_1) - v_0(\gamma(t_1)) = t_2 - t_1. \tag{6.16}$$

On the other hand, by using the usual convolution trick to regularize v_0, we have, for any $\xi \in \mathrm{AC}\,([0,1],\mathbb{R}^2)$ with $\xi(0) = \gamma(t_1)$ and $\xi(1) = \gamma(t_2)$,

$$\begin{aligned} \int_0^1 \frac{|\dot{\xi}(s)|}{a(\xi(s))}\, ds &\ge \int_0^1 \frac{\dot{\xi}(s) \cdot (p_0 + Dv_0(\xi(s)))}{a(\xi(s))|p_0 + Dv_0(\xi(s))|}\, ds \\ &= p_0 \cdot \gamma(t_2) + v_0(\gamma(t_2)) - p_0 \cdot \gamma(t_1) - v_0(\gamma(t_1)). \end{aligned}$$

This, together with (6.15) and (6.16), yields that

$$d(\gamma(t_1), \gamma(t_2)) = t_2 - t_1 = \int_{t_1}^{t_2} \frac{|\dot{\gamma}(s)|}{a(\gamma(s))}\, ds.$$

Thus, γ is a minimizing geodesic. $\square$

Let us now list various special properties in the two dimensional setting.

(1) Two minimizing geodesics intersect at most once.

(2) Periodic minimizing orbits of the same period are completely ordered in $\mathbb{R}^2$. Two periodic minimizing orbits of the same period are called neighboring if there does not exist any other periodic minimizing orbit between them.

(3) For any minimizing geodesic $c : \mathbb{R} \to \mathbb{R}^2$, the asymptotic slope

$$\alpha_c = \lim_{t\to\infty} \frac{c(t)}{|c(t)|} = \lim_{t\to-\infty} -\frac{c(t)}{|c(t)|}$$

exists.

(4) If α_c is rational, then c is either periodic or asymptotic to periodic minimizing geodesics as $t \to \pm\infty$. In the latter case, c could be a heteroclinic orbit connecting two periodic orbits for example.

(5) If α_c is irrational, then the set of minimizing geodesics with asymptotic slope α_c is ordered. In particular, two such minimizing geodesics do not cross.

Definition 6.15. Let c, c_1, c_2 be given minimizing geodesics. Denote by Γ_c^+ the region in $\mathbb{R}^2$ that stays above c, and Γ_c^- the region in $\mathbb{R}^2$ that stays below c.

We say that $c_1 < c_2$ if c_1 stays strictly below c_2 in $\mathbb{R}^2$, or equivalently, if c_2 stays in $\Gamma_{c_1}^+$.

Definition 6.16 (Function u_c)**.** Let $c : \mathbb{R} \to \mathbb{R}^2$ be a minimizing geodesic. We define a Lipschitz continuous function $u_c : \mathbb{R}^2 \to \mathbb{R}$ such that

$$u_c(x) = \lim_{t\to\infty} \left(d(x, c(t)) - d(c(t), c(0))\right). \tag{6.17}$$

If we normalize c such that $d(c(t), c(0)) = t$ for $t \geq 0$, then the above becomes

$$u_c(x) = \lim_{t\to\infty} \left(d(x, c(t)) - t\right).$$

Intuitively, we kick the vertex $c(t)$ of the cone $d(x, c(t))$ in (6.17) to infinity as $t \to \infty$.

Lemma 6.17. *Assume* (6.14)*. Then,* u_c *defined in* (6.17) *is well-defined and is a viscosity solution to*

$$a(x)|Du_c| = 1 \qquad \text{in } \mathbb{R}^2.$$

Proof. For $t \geq 0$, denote by

$$u^t(x) = d(x, c(t)) - d(c(t), c(0)) \qquad \text{for } x \in \mathbb{R}^2.$$

Clearly, by the triangle inequality, for fixed $x \in \mathbb{R}^2$,

$$u^t(x) = d(x, c(t)) - d(c(t), c(0)) \geq -d(x, c(0)).$$

For $0 \le t_1 \le t_2$, by the triangle inequality again,

$$\begin{aligned} u^{t_2}(x) - u^{t_1}(x) &= (d(x,c(t_2)) - d(c(t_2),c(0))) - (d(x,c(t_1)) - d(c(t_1),c(0))) \\ &= d(x,c(t_2)) - d(x,c(t_1)) - d(c(t_1),c(t_2)) \le 0. \end{aligned}$$

As $t \mapsto u^t(x)$ is nonincreasing and bounded from below, we imply that u_c is well-defined.

Besides, u^t is a viscosity solution to

$$\begin{cases} a(x)|Du^t(x)| = 1 & \text{in } \mathbb{R}^2 \setminus \{c(t)\}, \\ u^t(c(t)) = -d(c(t),c(0)). \end{cases}$$

Let $t \to \infty$ and use the stability result for viscosity solutions to conclude. □

We note further that

$$u_c(c(s)) = -d(c(s),c(0))\operatorname{sign}(s) \qquad \text{for } s \in \mathbb{R}.$$

After a suitable normalization, c is a global characteristic of u_c.

Definition 6.18 (Co-rays)**.** Let c be a minimizing geodesic. Let u_c be defined as in (6.17). A ray $\tilde{c} : [0,\infty) \to \mathbb{R}^2$ is called a co-ray of c if, for all $s \ge 0$,

$$u_c(\tilde{c}(s)) = u_c(\tilde{c}(0)) - d(\tilde{c}(s),\tilde{c}(0)).$$

It is important noting that if $\tilde{c}$ is a co-ray of c, then u_c is differentiable at $\tilde{c}(s)$ for $s > 0$, and

$$Du_c(\tilde{c}(s)) = -\frac{\dot{\tilde{c}}(s)}{a(\tilde{c}(s))|\tilde{c}(s)|}. \tag{6.18}$$

The lemma below gives a simple way to construct co-rays.

Lemma 6.19. *Assume* (6.14)*. Let c be a minimizing geodesic. Let u_c be defined as in* (6.17)*. Fix $x \in \mathbb{R}^2$. For $k \in \mathbb{N}$, let $\gamma_k : [0,k] \to \mathbb{R}^2$ be a minimizing geodesic connecting x to $c(k)$, that is, $\gamma_k(0) = x$, and $\gamma_k(k) = c(k)$. Assume that there exists a subsequence $\{\gamma_{k_j}\}$ of $\{\gamma_k\}$ that converges locally uniformly to $\gamma \in \mathrm{AC}([0,\infty),\mathbb{R}^2)$. Then, γ is a co-ray of c.*

Proof. By stability results, we have first that γ is a ray.

We next prove that γ is a co-ray of c. For any $s \ge 0$,

$$u_c(\gamma(0)) - u_c(\gamma(s)) \le d(\gamma(0),\gamma(s)).$$

We now prove the reverse inequality. For any $k \in \mathbb{N}$ with $k > s$, it is clear that

$$\begin{aligned} u_c(\gamma_k(s)) &\le u_c(c(k)) + d(\gamma_k(s),c(k)) \\ &= -d(c(0),c(k)) + d(x,c(k)) - d(x,\gamma_k(s)). \end{aligned}$$

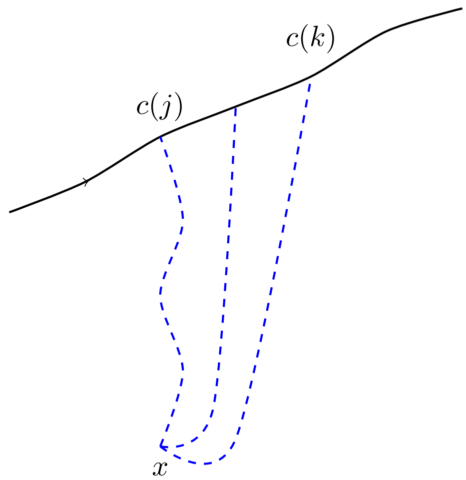

Figure 2. Construction of a co-ray

By the definition of u_c,

$$\lim_{k\to\infty}(-d(c(0),c(k))+d(x,c(k)))=u_c(x).$$

Combine the two relations above and let $k\to\infty$ to imply

$$u_c(\gamma(s))\le u_c(\gamma(0))-d(\gamma(0),\gamma(s)).$$

Thus,

$$u_c(\gamma(s))=u_c(\gamma(0))-d(\gamma(0),\gamma(s)).$$

The proof is complete. □

We next have the following property of co-rays.

Theorem 6.20. *Assume* (6.14). *Let c be a periodic minimizing geodesic. Let u_c be defined as in* (6.17). *Then, a ray $\tilde{c}:[0,\infty)\to\mathbb{R}^2$ is a co-ray of c if one of the following holds.*

(1) *$\tilde{c}$ is a part of a periodic minimizing geodesic.*
(2) *There exists a periodic minimizing geodesic c^+ such that $c^+>c$, and $\tilde{c}$ is asymptotic to c^+ from above as $t\to\infty$.*
(3) *There exists a periodic minimizing geodesic c^- such that $c^-<c$, and $\tilde{c}$ is asymptotic to c^- from below as $t\to\infty$.*

Lemma 6.21. *Assume* (6.14). *Let c be a periodic minimizing geodesic with asymptotic slope α. Let u_c be defined as in* (6.17). *Then, there exist*

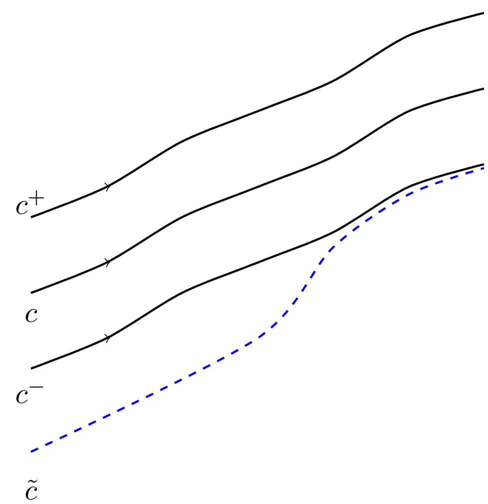

Figure 3. Characterizations of co-rays

$p^+, p^- \in \mathbb{R}^2$ *and periodic functions* v^+, v^- *such that*

$$\begin{cases} u_c(x) - p^+ \cdot x = v^+(x) & \text{for } x \in \Gamma_c^+, \\ u_c(x) - p^- \cdot x = v^-(x) & \text{for } x \in \Gamma_c^-. \end{cases}$$

Proof. We just need to prove the existence of p^- as the proof of the existence of p^+ is analogous. For any $v \in \mathbb{Z}^2$, consider $\Omega_v = \Gamma_c^- \cap \Gamma_{c-v}^-$.

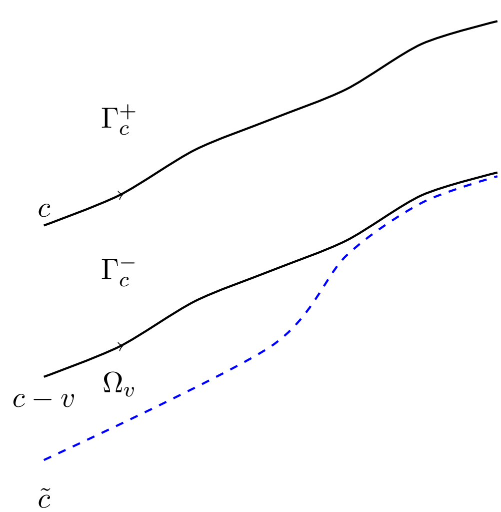

Figure 4. Common co-ray $\tilde{c}$ of c and $c - v$

Surely,

$$u_c(x+v) - u_c(x) = u_{c-v}(x) - u_c(x).$$

Take $x \in \Omega_v$ to be a point of differentiability of both u_c and u_{c-v}. Thanks to Lemma 6.19 and Theorem 6.20, c and $c-v$ have a common co-ray $\tilde{c}$ starting from x. We then use (6.18) to get that

$$Du_{c-v}(x) = Du_c(x).$$

Therefore, $u_{c-v} - u_c$ is constant in Ω_v. Hence, there exists a constant $l(v) \in \mathbb{R}$ such that

$$u_c(x+v) - u_c(x) = u_{c-v}(x) - u_c(x) = l(v) \qquad \text{for } x \in \Omega_v.$$

By shifting, it is clear that $l(v)$ is linear in v, which yields the existence of p^- immediately. $\square$

Suppose that c and $\tilde{c}$ are two neighboring periodic minimizing geodesics. Let $\gamma^\pm$ be two heteroclinic orbits between c and $\tilde{c}$. By suitable parametrizations, we may assume that

$$\lim_{t\to\infty} |\gamma^+(t) - c(t)| = \lim_{t\to\infty} |\gamma^-(t) - \tilde{c}(t)| = 0,$$
$$\lim_{t\to-\infty} |\gamma^+(t) - \tilde{c}(t)| = \lim_{t\to-\infty} |\gamma^-(t) - c(t)| = 0.$$

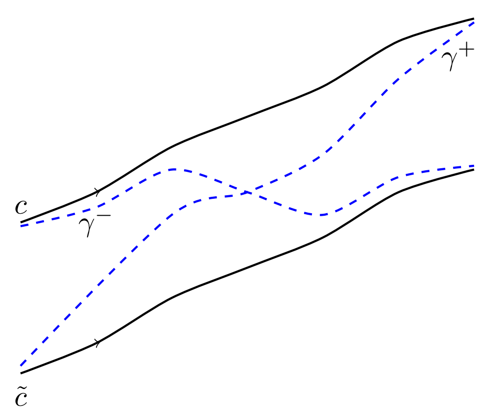

Figure 5. Heteroclinic orbits $\gamma^\pm$

Since γ^+ and γ^- cross, we have

$$b(c,\tilde{c}) = \lim_{t\to\infty} \Big[d(\gamma^+(t), \gamma^+(-t)) + d(\gamma^-(t), \gamma^-(-t)) - d(c(t), c(-t)) - d(\tilde{c}(t), \tilde{c}(-t)) \Big] > 0.$$

As $c, \tilde{c}$ are periodic, by shifting $c(0), \tilde{c}(0)$ to the left, we see that

$$0 < b(c,\tilde{c}) = u_c(\tilde{c}(0)) + u_{\tilde{c}}(c(0)). \tag{6.19}$$

Lemma 6.22. *Assume (6.14). Assume that c_1, c_2, c_3 are three periodic minimizing geodesics of the same period with $c_1 < c_2 < c_3$. Then,*

$$b(c_1, c_3) = b(c_1, c_2) + b(c_2, c_3).$$

Proof. It is clear that

$$u_{c_1}(c_3(0)) = u_{c_2}(c_3(0)) + u_{c_1}(c_2(0)),$$

and

$$u_{c_3}(c_1(0)) = u_{c_2}(c_1(0)) + u_{c_3}(c_2(0)).$$

Combine the two identities to conclude. $\square$

Lemma 6.23. *Assume the settings in Lemma 6.21. Assume further that $p^+ = p^-$. Then, $\mathbb{R}^2$ is foliated by minimizing periodic orbits of the same asymptotic slope α. In other words, for any $x \in \mathbb{R}^2$, there exists a periodic minimizing geodesic with asymptotic slope α passing through x.*

Proof. It suffices to show that there do not exist two neighboring minimizing periodic orbits of the same asymptotic slope α.

Let c be a periodic minimizing geodesic with asymptotic slope α as given in Lemma 6.21. For $v = (-n, m) \in \mathbb{Z}^2$ with $n > 0$, denote by $c^+ = c + v$ and $c^- = c - v$. Then,

$$c^- < c < c^+.$$

As $u_c(c(0)) = 0$, we have that

$$\begin{cases} u_c(c(0) - v) = u_c(c(0) - v) - u_c(c(0)) = -p^- \cdot v, \\ u_{c-v}(c(0)) = u_{c-v}(c(0)) - u_c(c(0)) = u_c(c(0) + v) - u_c(c(0)) = p^+ \cdot v. \end{cases}$$

Since $p^- = p^+$, we imply that

$$u_c(c(0) - v) + u_{c-v}(c(0)) = 0,$$

which means that

$$b(c, c - v) = 0. \tag{6.20}$$

Combine this with (6.19) and Lemma 6.22, we yield that $c - v$ does not have any neighboring periodic orbit. The proof is complete. $\square$

6.4.2. Proof of Theorem 6.10.

Proof of Theorem 6.10. Assume that the level curve $\mathcal{C} = \{\overline{H} = 1\}$ is strictly convex. By Lemma 6.11, this level curve is also C^1. Hence, the map $G : \mathcal{C} \to S^1$ defined as

$$G(p) = \frac{D\overline{H}(p)}{|D\overline{H}(p)|} \qquad \text{for } p \in \mathcal{C}$$

is continuous and one-to-one.

For any rational vector $\alpha \in S^1$, Let c be a periodic minimizing geodesic with asymptotic slope α. Let $p^{\pm}$ be as in Lemma 6.21. By Lemma 6.14, we imply that there exist periodic functions $v^+, v^- \in C(\mathbb{T}^2)$ such that

$$a(y)|p^+ + Dv^+(y)| = a(y)|p^- + Dv^-(y)| = 1 \qquad \text{in } \mathbb{T}^2.$$

Thus,

$$\overline{H}(p^+) = \overline{H}(p^-) = 1 \qquad \text{and} \qquad G(p^+) = G(p^-) = \alpha.$$

Therefore, $p^+ = p^-$. We then use Lemma 6.23 to deduce further that for any $x \in \mathbb{R}^2$ and any rational vector $\alpha \in S^1$, there exists a periodic minimizing geodesic with asymptotic slope α passing through x.

By approximations and stability, we derive that for any $x \in \mathbb{R}^2$ and any vector $\alpha \in S^1$, there exists a minimizing geodesic $\gamma_{x,\alpha}$ with asymptotic slope α passing through x. Of course, $\gamma_{x,\alpha}$ is unique. It is then not hard to see that for fixed x, the map $\alpha \mapsto \frac{\dot{\gamma}_{x,\alpha}}{|\dot{\gamma}_{x,\alpha}|}$ is continuous and one-to-one. The map is hence onto as well. Therefore, any geodesic is a minimizing geodesic and there does not exist conjugate points. This implies that the metric is flat. $\square$

6.5. References

(1) Much of the content in this chapter is based on Bangert [**Ban88**].

(2) Theorem 6.6 was obtained by Carneiro in [**Car95**].

(3) Theorem 6.10 was proved by Bangert [**Ban94**].

Chapter 7

Aubry-Mather theory in the merely continuous setting

In this chapter, we are always in the merely continuous setting. The main focus will be in two dimensions.

In the first section, we give some basic results of the weak KAM theory for the classical mechanic Hamiltonian in the merely continuous setting. Because of the lack of smoothness, we need to proceed with care.

7.1. Classical mechanic Hamiltonian

We assume throughout this section

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

where $V \in C(\mathbb{T}^n)$. Then, the cell problem reads

$$\frac{1}{2}|p + Dv(y)|^2 + V(y) = \overline{H}(p) \qquad \text{in } \mathbb{T}^n. \tag{7.1}$$

It is important to note that we only have V is merely continuous on $\mathbb{T}^n$.

7.1.1. Differentiability property. The following lemma plays an important role in this setting.

Lemma 7.1. *Let U be an open subset of $\mathbb{R}^n$ and $c \in \mathbb{R}$. For $i = 1, 2$, assume $w_i \in W^{1,\infty}(U)$ satisfies*

$$\frac{1}{2}|Dw_i|^2 + V(x) \le c \qquad \text{for a.e. } x \in U.$$

Assume further that there exists a curve $\eta \in \mathrm{AC}([a,b],U)$ such that, for $i=1,2$,

$$\int_a^b \left(\frac{1}{2}|\dot\eta(t)|^2 - V(\eta(t)) + c\right) dt = w_i(\eta(b)) - w_i(\eta(a)).$$

Then, the following properties hold.

(1) *For a.e. $t \in [a,b]$,*

$$\frac{1}{2}|\dot\eta(t)|^2 + V(\eta(t)) = c. \tag{7.2}$$

(2) *If η is differentiable at $t_0 \in (a,b)$, then both w_1 and w_2 are differentiable at $x=\eta(t_0)$, and*

$$Dw_1(\eta(t_0)) = Dw_2(\eta(t_0)) = \dot\eta(t_0).$$

(3) *For all $t \in (a,b)$, $w_1 - w_2$ is differentiable at $x = \eta(t)$, and*

$$D(w_1-w_2)(x) = 0.$$

Proof. By translations, we assume that $0 \in (a,b)$ and $\eta(0)=0$.

We first prove (7.2). As usual, by standard mollification of w_1 and approximations, we have that

$$\begin{aligned} w_1(\eta(b)) - w_1(\eta(a)) &= \int_a^b p_1(t)\cdot\dot\eta(t)\,dt \\ &= \int_a^b \frac{1}{2}\left(|p_1(t)|^2 + |\dot\eta(t)|^2 - |p_1(t)-\dot\eta(t)|^2\right) dt \end{aligned}$$

for some $p_1(t) \in \partial w_1(\eta(t))$ for $t\in(a,b)$. Here,

$$\partial w_1(x) = \mathrm{co}(K(x)),$$

where $\mathrm{co}(K(x))$ is the convex hull of the set

$$K(x) = \left\{p \in \mathbb{R}^n \,:\, \exists\,\{x_k\} \to x \text{ s.t. } Dw_1(x_k) \text{ exists, } p = \lim_{k\to\infty} Dw_1(x_k)\right\}.$$

Also, through the approximation process,

$$\frac{1}{2}|p_1(t)|^2 + V(\eta(t)) \le c \qquad \text{for all } t \in [a,b].$$

Combining the above points with the hypothesis, we yield

$$\frac{1}{2}|p_1(t)|^2 + V(\eta(t)) = c \qquad \text{and} \qquad p_1(t) = \dot\eta(t) \quad \text{for a.e. } t\in[a,b].$$

Hence (7.2) holds. Moreover, we get that η is Lipschitz continuous.

Let us now prove (2) and (3). It suffices to show that for a sequence $\{\lambda_m\} \subset (0,\infty)$ with $\lim_{m\to\infty}\lambda_m = 0$, if

$$\lim_{m\to\infty}\frac{\eta(\lambda_m t)}{\lambda_m} \qquad \text{exists for all } t\in\mathbb{R},$$

then, for all $t \in \mathbb{R}$, $x \in \mathbb{R}^n$, and $i = 1, 2$,

$$\lim_{m\to\infty} \frac{\eta(\lambda_m t)}{\lambda_m} = qt \qquad \text{and} \qquad \lim_{m\to\infty} \frac{w_i(\lambda_m x) - w_i(0)}{\lambda_m} = q \cdot x$$

for some q satisfying $\frac{1}{2}|q|^2 + V(0) = c$. It is enough to prove this claim for w_1 as the proof for w_2 is analogous. Let

$$\bar{\eta}(t) = \lim_{m\to\infty} \frac{\eta(\lambda_m t)}{\lambda_m} \qquad \text{for } t \in \mathbb{R}.$$

By the hypothesis, η is an absolute minimizer of the action

$$\int (\frac{1}{2}|\dot{\gamma}(t)|^2 - V(\gamma(t)) + c)\, dt$$

on $[a, b]$. Then, $\eta(\lambda_m t)/\lambda_m$ is an absolute minimizer of the action

$$\int \left(\frac{1}{2}|\dot{\gamma}(t)|^2 - V(\lambda_m \gamma(t)) + c\right) dt$$

on $[a/\lambda_m, b/\lambda_m]$. By the stability of minimizing curves, $\bar{\eta}$ is an absolute minimizer of the action

$$\int \left(\frac{1}{2}|\dot{\gamma}(t)|^2 - V(0) + c\right) dt,$$

and $\bar{\eta}(0) = 0$. By the Euler-Lagrange equations,

$$\ddot{\bar{\eta}}(t) = 0 \qquad \text{for } t \in \mathbb{R},$$

and hence

$$\bar{\eta}(t) = tq$$

for some $q \in \mathbb{R}^n$. In light of (7.2),

$$|q| \le M = \sqrt{2(c - V(0))}.$$

By passing to a subsequence if necessary, we assume

$$\lim_{m\to\infty} \frac{w_1(\lambda_m x) - w_1(0)}{\lambda_m} = u(x) \qquad \text{for all } x \in \mathbb{R}^n.$$

Then, $u \in W^{1,\infty}(\mathbb{R}^n)$, and

$$\frac{1}{2}|Du(x)|^2 + V(0) \le c \qquad \text{for a.e. } x \in \mathbb{R}^n,$$

which is equivalent to

$$|Du(x)| \le M \qquad \text{for a.e. } x \in \mathbb{R}^n.$$

Thanks to (7.2),

$$\begin{aligned} w_1(\eta(\lambda_m t)) - w_1(\eta(0)) &= \int_0^{\lambda_m t} \left(\frac{1}{2}|\dot\eta(s)|^2 - V(\eta(s)) + c\right) ds \\ &= \int_0^{\lambda_m t} 2\left(c - V(\eta(s))\right) ds. \end{aligned}$$

Dividing both sides by λ_m, and sending $m \to \infty$, we imply

$$u(qt) = tM^2 \qquad \text{for all } t \in \mathbb{R}.$$

As $|Du| \le M$,

$$u(qt) = u(qt) - u(0) \le M|q|t \le M^2 t.$$

Therefore, we deduce that $|q| = M$, and

$$u(et) = Mt \qquad \text{for } t \in \mathbb{R}$$

for $e = \frac{q}{|q|} = \frac{q}{M}$. We claim that

$$u(x) = q \cdot x \qquad \text{for } x \in \mathbb{R}^n. \tag{7.3}$$

Indeed, as $|Du| \le M$,

$$|u(x) - Mt| = |u(x) - u(et)| \le M|x - et| \qquad \text{for } x \in \mathbb{R}^n,\ t \in \mathbb{R}.$$

Taking square of both sides to get

$$u(x)^2 - 2Mtu(x) + M^2t^2 \le M^2(|x|^2 - 2x \cdot et + t^2),$$

which is reduced to

$$u(x)^2 - M^2|x|^2 \le 2Mt(u(x) - q \cdot x).$$

Let $t \to \pm\infty$ in the above to conclude that

$$u(x) = q \cdot x.$$

□

7.1.2. The Aubry set. Let us now proceed to define the Aubry set, which is analogous to the smooth setting. Nevertheless, we recall everything here for clarity. Instead of phrasing everything on the torus $\mathbb{T}^n$, we lift all the curves to $\mathbb{R}^n$. For $t > 0$, and $p, x, y \in \mathbb{R}^n$, denote by

$$G_{t,p}(x,y) = \inf_{\substack{\xi \in \mathrm{AC}\,([0,t],\mathbb{R}^n),\\ \xi(0)=x,\ \xi(t) \in y + \mathbb{Z}^n}} \int_0^t \left(\frac{1}{2}|\dot\xi(s)|^2 - V(\xi(s)) - p \cdot \dot\xi(s) + \overline{H}(p)\right) ds,$$

and

$$G_p(x,y) = \liminf_{t \to \infty} G_{t,p}(x,y).$$

By Lemma 3.14, $G_{t,p}(x,y) \ge v(y) - v(x)$, and hence

$$G_p(x,y) \ge v(y) - v(x)$$

for any viscosity solution $v \in C(\mathbb{T}^n)$ of (7.1). In particular,

$$G_p(x,x) \geq 0.$$

Besides, $G_p(x,y)$ is $\mathbb{Z}^n$-periodic and Lipschitz continuous in both x and y.

Definition 7.2. The Aubry set associated with $p \in \mathbb{R}^n$ is defined as

$$\mathcal{A}_p = \{x \in \mathbb{R}^n \,:\, G_p(x,x) = 0\}.$$

Lemma 7.3. *Let $v \in C(\mathbb{T}^n)$ be a viscosity solution of (7.1) for some $p \in \mathbb{R}^n$ fixed. Suppose that $\{\xi_m\}$ is a sequence of global characteristics of v such that*

$$\lim_{m\to\infty} \xi_m = \xi \qquad \text{locally uniformly in } \mathbb{R}.$$

Then, ξ is also a global characteristic of v.

Proof. Denote by $u(x) = p \cdot x + v(x)$ for $x \in \mathbb{R}^n$. Fix $t_1 < t_2$. For $m \in \mathbb{N}$, we have that

$$u(\xi_m(t_2)) - u(\xi_m(t_1)) = \int_{t_1}^{t_2} \left(\frac{1}{2}|\dot{\xi}_m(s)|^2 - V(\xi(s)) + \overline{H}(p)\right) ds.$$

Sending $m \to \infty$ and using the weakly lower semicontinuity of the integral to imply

$$\begin{aligned} u(\xi(t_2)) - u(\xi(t_1)) &\geq \liminf_{m\to\infty} \int_{t_1}^{t_2} \left(\frac{1}{2}|\dot{\xi}_m(s)|^2 - V(\xi_m(s)) + \overline{H}(p)\right) ds \\ &\geq \int_{t_1}^{t_2} \left(\frac{1}{2}|\dot{\xi}(s)|^2 - V(\xi(s)) + \overline{H}(p)\right) ds. \end{aligned}$$

Combining this with Lemma 3.14, we get the conclusion. □

Definition 7.4. Fix $p \in \mathbb{R}^n$. A Lipschitz continuous curve $\xi : \mathbb{R} \to \mathbb{R}^n$ is called a universal global characteristic associated with p if it is a global characteristic of every viscosity solution v of the cell problem (7.1).

Denote by $\mathcal{U}_p$ the collection of all universal characteristics associated with p. It is then clear that, for every $\xi \in \mathcal{U}_p$,

$$\frac{1}{2}|\dot{\xi}(s)|^2 + V(\xi(s)) = \overline{H}(p) \qquad \text{for a.e. } t \in \mathbb{R}. \tag{7.4}$$

Besides, $\mathcal{U}_p$ is closed in the locally uniform topology thanks to Lemma 7.3.

Lemma 7.5. *For every $p \in \mathbb{R}^n$,*

$$\mathcal{A}_p \neq \emptyset.$$

We have already proved this lemma in Theorem 5.20 in the smooth setting. The merely continuous setting then follows by suitable approximations. Nevertheless, it is natural to give a direct proof here.

Proof. Let v be a viscosity solution of (7.1). Let $\xi : \mathbb{R} \to \mathbb{R}^n$ be a global characteristic of with v. By projecting ξ back to $\mathbb{T}^n$ and using a suitable translation in time, we find two sequences $\{t_m\} \to +\infty$ and $\{x_m\} \subset [0,1]^n$ such that

$$\begin{cases} \lim_{m\to\infty}(t_{m+1}-t_m) = +\infty, \\ \xi(t_m) = x_m + k_m \qquad \text{for some } k_m \in \mathbb{Z}^n, \\ \lim_{m\to\infty} x_m = x_0 \in [0,1]^n, \end{cases}$$

and

$$v(x_{m+1}) - v(x_m) = \int_{t_m}^{t_{m+1}} \left(\frac{1}{2}|\dot{\xi}(s)|^2 - V(\xi(s)) - p\cdot\dot{\xi}(s) + \overline{H}(p)\right) ds.$$

Then,

$$G_{t_{m+1}-t_m,p}(x_m, x_{m+1}) = v(x_{m+1}) - v(x_m).$$

Let $m \to \infty$ to deduce that

$$G_p(x_0, x_0) \le \liminf_{m\to\infty} G_{t_{m+1}-t_m,p}(x_m, y_m) = 0,$$

which means $x_0 \in \mathcal{A}_p$.

$\square$

Lemma 7.6. *For any $x \in \mathcal{A}_p$, there exists $\xi \in \mathcal{U}_p$ such that $\xi(0) = x$. In particular, $\mathcal{U}_p \neq \emptyset$.*

Proof. Fix $x \in \mathcal{A}_p$. By the definition of $\mathcal{A}_p$, there exist $\{t_m\} \to \infty$ and a sequence of curves $\gamma_m : [0, t_m] \to \mathbb{R}^n$ such that $\gamma_m(0) = x$, and $\gamma_m(t_m) = x + k_m$ for some $k_m \in \mathbb{Z}^n$, and

$$\lim_{m\to\infty} \int_0^{t_m} \left(\frac{1}{2}|\dot{\gamma}_m(s)|^2 - V(\gamma_m(s)) - p\cdot\dot{\gamma}_m(s) + \overline{H}(p)\right) ds = 0.$$

Let v be a viscosity solution of (7.1). Then, for $L > 0$ and m large enough,

$$v(\gamma_m(L)) - v(x) \le \int_0^L \left(\frac{1}{2}|\dot{\gamma}_m(s)|^2 - V(\gamma_m(s)) - p\cdot\dot{\gamma}_m(s) + \overline{H}(p)\right) ds$$

and

$$v(\gamma_m(t_m)) - v(\gamma_m(L)) \le \int_L^{t_m} \left(\frac{1}{2}|\dot{\gamma}_m|^2 - V(\gamma_m) - p\cdot\dot{\gamma}_m + \overline{H}(p)\right) ds.$$

We use the above two equalities and the fact that

$$v(\gamma_m(L)) - v(x) + v(\gamma_m(t_m)) - v(\gamma_m(L)) = 0$$

to yield

$$\lim_{m\to\infty} \Big(\int_0^L \left(\frac{1}{2}|\dot{\gamma}_m(s)|^2 - V(\gamma_m(s)) - p\cdot\dot{\gamma}_m(s) + \overline{H}(p)\right) ds \\ - \left(v(\gamma_m(L)) - v(x)\right)\Big) = 0.$$

By using a similar logic,

$$\lim_{m\to\infty}\Big(\int_{t_m-L}^{t_m}\left(\frac{1}{2}|\dot\gamma_m(s)|^2-V(\gamma_m(s))-p\cdot\dot\gamma_m(s)+\overline{H}(p)\right)ds - \left(v(\gamma_m(t_m))-v(\gamma(t_m-L))\right)\Big)=0.$$

Define $\xi_m : [-t_m/2, t_m/2] \to \mathbb{R}^n$ as

$$\xi_m(t)=\begin{cases}\gamma_m(t) & \text{for } t\in\left[0,\frac{t_m}{2}\right],\\ \gamma_m(t_m+t)-k_m & \text{for } t\in\left[-\frac{t_m}{2},0\right]\end{cases}$$

Clearly, for any fixed $L>0$, $\{\|\xi_m\|_{H^1((-L,L))}\}$ is uniformly bounded. By passing to a subsequence if needed, we assume that

$$\lim_{m\to\infty}\xi_m(t)=\xi(t) \qquad \text{locally uniformly in } \mathbb{R}.$$

for $\xi\in \mathrm{AC}(\mathbb{R},\mathbb{R}^n)$. By the above points,

$$\lim_{m\to\infty}\Big(\int_{-L}^{L}\left(\frac{1}{2}|\dot\xi_m(s)|^2-V(\xi_m(s))-p\cdot\dot\xi_m(s)+\overline{H}(p)\right)ds - \left(v(\xi_m(L))-v(\xi_m(-L))\right)\Big)=0.$$

Thanks to the weakly lower semicontinuity of the integral, we see that, for any $L>0$,

$$\int_{-L}^{L}\left(\frac{1}{2}|\dot\xi(s)|^2-V(\xi(s))-p\cdot\dot\xi(s)+\overline{H}(p)\right)ds=v(\xi(L))-v(\xi(-L)).$$

Hence ξ is a universal global characteristic associated with p. $\square$

7.1.3. Modifications of global characteristics. For smooth V, two different orbits in the same Aubry set cannot intersect and two different absolute minimizers of the same action cannot intersect twice as discussed in the previous chapter. However, both situations could happen with merely continuous V. A key difference is that we do not have the corresponding Euler-Lagrange equations and the uniqueness property of their solutions in this merely continuous setting. Consequently, the structure of orbits on $\mathcal{U}_p$ might be very complicated, and the orbits might have pathological behaviors. It is then extremely hard to understand the topology of $\mathcal{U}_p$.

We now provide two procedures to join different pieces of two global characteristics, which will be used later to select nice minimizing orbits and then simplify the topology of interacting curves.

Definition 7.7 (Procedure 1)**.** Fix $p\in\mathbb{R}^n$. Let v be a viscosity solution of (7.1). Assume that ξ_1 and ξ_2 are two global characteristics of with v.

Assume further that for some $t_1, t_2 \in \mathbb{R}$,

$$\xi_1(t_1) = \xi_2(t_2).$$

We now glue a part of ξ_1 with a part of ξ_2 to have a new global characteristic of with v. Define

$$\xi_3(t) = \begin{cases} \xi_1(t) & \text{for } t \leq t_1 \\ \xi_2(t - t_1 + t_2) & \text{for } t \geq t_1. \end{cases}$$

See Figure 1.

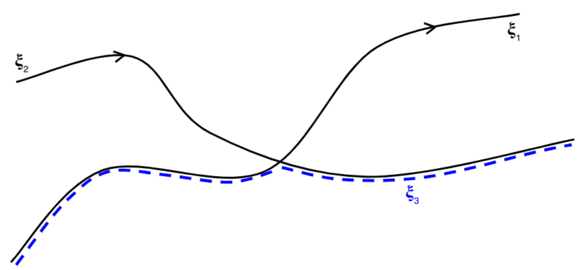

Figure 1. Formation of the curve ξ_3

Let us prove quickly that ξ_3 is also a global characteristic of with v. Indeed, for $a \leq t_1 \leq b$,

$$v(\xi_3(b)) - v(\xi_3(t_1)) = \int_{t_1}^{b} \left(\frac{1}{2}|\dot{\xi}_3(s)|^2 - V(\xi_3(s)) - p \cdot \dot{\xi}_3(s) + \overline{H}(p) \right) ds$$

and

$$v(\xi_3(t_1)) - v(\xi_3(a)) = \int_{a}^{t_1} \left(\frac{1}{2}|\dot{\xi}_3(s)|^2 - V(\xi_3(s)) - p \cdot \dot{\xi}_3(s) + \overline{H}(p) \right) ds.$$

Thus

$$\begin{aligned} v(\xi_3(b)) - v(\xi_3(a)) &= v(\xi_3(b)) - v(\xi_3(t_1)) + v(\xi_3(t_1)) - v(\xi_3(a)) \\ &= \int_{a}^{b} \left(\frac{1}{2}|\dot{\xi}_3(s)|^2 - V(\xi_3(s)) - p \cdot \dot{\xi}_3(s) + \overline{H}(p) \right) ds. \end{aligned}$$

Definition 7.8 (Procedure 2 – Crossing of two universal global characteristics)**.** Let $p, p' \in F_c = \{p \in \mathbb{R}^2 : \overline{H}(p) = c\}$. Let ξ and $\tilde{\xi}$ be orbits in $\mathcal{U}_p$ and $\mathcal{U}_{p'}$, respectively. Assume that there exist $t_1, t_2, t'_1, t'_2 \in \mathbb{R}$ such that, for $i = 1, 2$,

$$P_i = \xi(t_i) = \tilde{\xi}(t'_i).$$

We construct new orbits on $\mathcal{U}_p$ and $\mathcal{U}_{p'}$ by joining different pieces of ξ and $\tilde{\xi}$ in a very natural way. Without loss of generality, we assume that $t_1 < t_2$. There are two cases to be considered as following.

Case 1. $t_1' < t_2'$. Define

$$\xi_2(t) = \begin{cases} \xi(t) & \text{for } t \le t_1, \\ \tilde{\xi}(t + t_1' - t_1) & \text{for } t_1 \le t \le t_1 + t_2' - t_1', \\ \xi(t + t_2 - t_1 - t_2' + t_1') & \text{for } t_1 + t_2' - t_1' \le t. \end{cases}$$

See Figure 2.

Case 2. $t_1' > t_2'$. Define

$$\xi_3(t) = \begin{cases} \xi(t) & \text{for } t \le t_1, \\ \tilde{\xi}(t_1 + t_1' - t) & \text{for } t_1 \le t \le t_1 + t_1' - t_2', \\ \xi(t + t_2 - t_1 - t_1' + t_2') & \text{for } t_1 + t_1' - t_2' \le t. \end{cases}$$

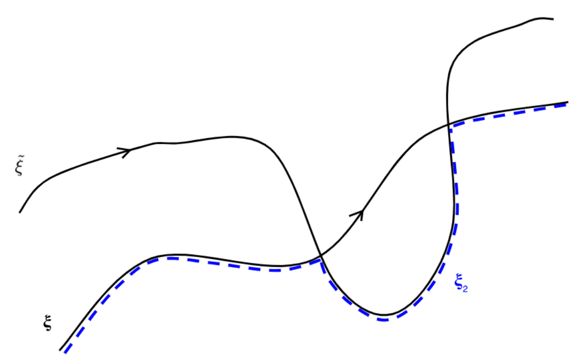

Figure 2. Combining two universal global characteristics

We say that ξ_2 or ξ_3 is the adjustment of ξ with respect to $\tilde{\xi}$ between t_1 and t_2. It is not hard to see that ξ_2 or ξ_3 belongs to $\mathcal{U}_p$.

Let us now state a relevant result on rotation vectors. We have already given a proof of a similar version of this in Theorem 3.15, and thus, we omit the proof here.

Lemma 7.9. *Let $p \in \mathbb{R}^n$, and v be a viscosity solution of* (7.1)*. Let $\xi : \mathbb{R} \to \mathbb{R}^n$ be a global characteristic of v. Assume that there exists $\{t_m\}$ converging to either $-\infty$ or $+\infty$ such that $\xi(t_m)/t_m$ converges as $m \to \infty$. Then,*

$$\lim_{m\to\infty} \frac{\xi(t_m)}{t_m} \in \partial \overline{H}(p).$$

Definition 7.10. Fix $p \in \mathbb{R}^n$. Let v be a viscosity solution of (7.1). Let $\xi : \mathbb{R} \to \mathbb{R}^n$ be a global characteristic of v. Then, ξ is called periodic if there exist $T > 0$ and $q \in \mathbb{Z}^n$ such that

$$\xi(t+T) - \xi(t) = q \quad \text{for all } t \in \mathbb{R}.$$

In this case, q/T is called the rotation vector of ξ.

Thanks to Lemma 7.9, the rotation vector

$$\frac{q}{T} \in \partial \overline{H}(p).$$

Also, it is clear that every periodic global characteristic associated with some v must be a universal global characteristic.

The following corollary follows rather straightforwardly from the above definition and Lemma 7.9.

Corollary 7.11. *Fix $p \in \mathbb{R}^n$. Let v be a viscosity solution of* (7.1). *Let $\xi : \mathbb{R} \to \mathbb{R}^n$ be a global characteristic of v. If there exist $t_1 < t_2$ such that*

$$\xi(t_2) - \xi(t_1) = q \in \mathbb{Z}^n,$$

then

$$\frac{q}{t_2 - t_1} \in \partial \overline{H}(p).$$

The following result is important and always needed in the construction of circle maps. The idea of the proof is similar to that of Proposition 6.5 and hence is omitted for now. Note that we do not need the smoothness assumption here as all can be done by approximations.

Lemma 7.12. *Assume that $n = 2$. For every $q \in \mathbb{Z}^2$ and $c > \max_{\mathbb{T}^2} V$, there exists $p_q \in F_c$ such that $\mathcal{U}_{p_q}$ has a periodic orbit ξ such that, for some $T > 0$,*

$$\xi(t+T) - \xi(t) = q \quad \textit{for all } t \in \mathbb{R}.$$

Next, we have the following lemma on the strict convexity of $\overline{H}$.

Lemma 7.13. *Suppose that there exist $p_0, p_1 \in \mathbb{R}^n$ and $\lambda \in (0,1)$ such that, for $p_\lambda = \lambda p_0 + (1-\lambda) p_1$,*

$$\overline{H}(p_\lambda) = \lambda \overline{H}(p_0) + (1-\lambda)\overline{H}(p_1).$$

Then,

$$\begin{cases} \overline{H}(p_\lambda) = \overline{H}(p_0) = \overline{H}(p_1), \\ \mathcal{A}_{p_\lambda} \subset \mathcal{A}_{p_0} \cap \mathcal{A}_{p_1}, \\ \mathcal{U}_{p_\lambda} \cap \mathcal{U}_{p_0} \cap \mathcal{U}_{p_1} \neq \emptyset. \end{cases}$$

Proof. We divide the proof into two steps for clarity.

Step 1. For $x \in \mathcal{A}_{p_\lambda}$, there exist $\{t_m\} \to \infty$ and a sequence of curves $\gamma_m : [0, t_m] \to \mathbb{R}^n$ such that $\gamma_m(0) = x$, $\gamma_m(t_m) \in x + \mathbb{Z}^n$, and

$$\lim_{m\to\infty} \int_0^{t_m} \left(\frac{1}{2}|\dot\gamma_m(s)|^2 - V(\gamma_m(s)) - p_\lambda \cdot \dot\gamma_m(s) + \overline{H}(p_\lambda) \right) ds = 0.$$

Let

$$A_m = \lambda \int_0^{t_m} \left(\frac{1}{2}|\dot\gamma_m(s)|^2 - V(\gamma_m(s)) - p_0 \cdot \dot\gamma_m(s) + \overline{H}(p_0) \right) ds,$$

and

$$B_m = (1-\lambda) \int_0^{t_m} \left(\frac{1}{2}|\dot\gamma_m(s)|^2 - V(\gamma_m(s)) - p_1 \cdot \dot\gamma_m(s) + \overline{H}(p_1) \right) ds.$$

Then, as $p_\lambda = \lambda p_0 + (1-\lambda)p_1$,

$$\int_0^{t_m} \left(\frac{1}{2}|\dot\gamma_m(s)|^2 - V(\gamma_m(s)) - p_\lambda \cdot \dot\gamma_m(s) + \overline{H}(p_\lambda) \right) ds = A_m + B_m.$$

By Lemma 3.14,

$$A_m, B_m \geq 0.$$

As $\lim_{m\to\infty}(A_m + B_m) = 0$, we then deduce that

$$\lim_{m\to\infty} A_m = \lim_{m\to\infty} B_m = 0.$$

Meanwhile, by the definition of $G_p(x,x)$, it is clear that

$$\begin{cases} 0 \leq \lambda G_{p_0}(x,x) \leq \lim_{m\to\infty} A_m, \\ 0 \leq (1-\lambda) G_{p_1}(x,x) \leq \lim_{m\to\infty} B_m. \end{cases}$$

Therefore, $G_{p_0}(x,x) = G_{p_1}(x,x) = 0$, which means that $x \in \mathcal{A}_{p_0} \cap \mathcal{A}_{p_1}$. We thus have

$$\mathcal{A}_{p_\lambda} \subset \mathcal{A}_{p_0} \cap \mathcal{A}_{p_1}.$$

Moreover, by using the proof of Lemma 7.6, we introduce a suitable reparametrization of $\{\gamma_m\}$ to get a sequence of curves converging locally uniformly to a common orbit in $\mathcal{U}_{p_\lambda} \cap \mathcal{U}_{p_0} \cap \mathcal{U}_{p_1}$.

Step 2. Fix an orbit $\xi \in \mathcal{U}_{p_0} \cap \mathcal{U}_{p_1}$. Then, by (7.4),

$$\begin{cases} \frac{1}{2}|\dot\xi(t)|^2 + V(\xi(t)) = \overline{H}(p_0) & \text{for a.e } t \in \mathbb{R}, \\ \frac{1}{2}|\dot\xi(t)|^2 + V(\xi(t)) = \overline{H}(p_1) & \text{for a.e } t \in \mathbb{R}. \end{cases}$$

Therefore, $\overline{H}(p_0) = \overline{H}(p_1)$.

□

Thanks to Lemma 7.13, we have immediately the following result.

Theorem 7.14. *Assume that $H(y,p) = \frac{1}{2}|p|^2 + V(y)$ for $(y,p) \in \mathbb{T}^n \times \mathbb{R}^n$ for some $V \in C(\mathbb{T}^n)$. Fix $p_0, p_1 \in \mathbb{R}^n$. If $\overline{H}(p_0) \neq \overline{H}(p_1)$, then for all $\lambda \in (0,1)$,*

$$\overline{H}(\lambda p_1 + (1-\lambda)p_0) < \lambda \overline{H}(p_0) + (1-\lambda)\overline{H}(p_1),$$

that is, $\overline{H}$ is strictly convex along directions that are not tangential to each given level set.

7.2. The two dimensional setting

We are always in the two dimensional setting, that is, $n = 2$ in this section. Let us state the main theorem of this section right away.

Theorem 7.15. *Assume $n = 2$. Assume that $H(y,p) = \frac{1}{2}|p|^2 + V(y)$ for $(y,p) \in \mathbb{T}^2 \times \mathbb{R}^2$ for some $V \in C(\mathbb{T}^2)$. Then, for $n = 2$ and $c > \max_{\mathbb{T}^2} V$, $F_c = \left\{p \in \mathbb{R}^2 : \overline{H}(p) = c\right\}$ does not contain a line segment of irrational slope.*

The proof of this theorem is rather long, and we will proceed through various steps.

Proof. We give a proof by contradiction.

Step 1. Assume otherwise that F_c contains a line segment of an irrational slope. Let p_0 and p_1 be two points in the interior of this line segment. Thanks to Lemma 7.13,

$$\mathcal{A}_{p_0} = \mathcal{A}_{p_1},$$

and

$$p_0 - p_1 \text{ is an irrational vector.}$$

Then, the outward unit normal vector $\vec{n}$ is also irrational, and

$$\partial\overline{H}(p_0) = \partial\overline{H}(p_1) = \{\lambda\vec{n} \,:\, \lambda \in [\alpha,\beta]\}$$

for two positive numbers $0 < \alpha < \beta$. Without loss of generality, we assume that

$$\vec{n} \cdot (1,0) > 0. \tag{7.5}$$

Let v_0 and v_1 be viscosity solutions to (7.1) corresponding to $p = p_0$ and $p = p_1$, respectively. Denote by

$$\mathcal{U} = \mathcal{U}_{p_0} \cap \mathcal{U}_{p_1},$$

and

$$S = \bigcup_{\xi \in \mathcal{U}} \xi(\mathbb{R}) \subset \mathbb{R}^2.$$

In light of Lemma 7.13,

$$\mathcal{U} \neq \emptyset.$$

Step 2. For $x \in \mathbb{R}^2$, set

$$\begin{cases} u_0(x) = p_0 \cdot x + v_0(x), \\ u_1(x) = p_1 \cdot x + v_1(x). \end{cases}$$

By Lemma 7.1, $u_0 - u_1$ is differentiable at $x \in S$, and

$$D(u_0 - u_1)(x) = 0 \qquad \text{for } x \in S. \tag{7.6}$$

Note again that we do not know if u_0 or u_1 is differentiable at x individually yet. Thanks to Lemma 7.12, we pick $p' \in F_c$ such that $\mathcal{U}_{p'}$ contains a periodic orbit η such that for some $T > 0$

$$\eta(t+T) - \eta(t) = (0,1) = e_2 \qquad \text{for all } t \in \mathbb{R}.$$

For $e_1 = (1,0)$, denote

$$\Lambda = \max\{|e_1 \cdot (x-y)| \,:\, x, y \in \eta(\mathbb{R})\}. \tag{7.7}$$

Choose a positive integer $J > \Lambda + 1$ and for $k \in \mathbb{Z}$, denote

$$\eta_k = \eta + k(J, 0).$$

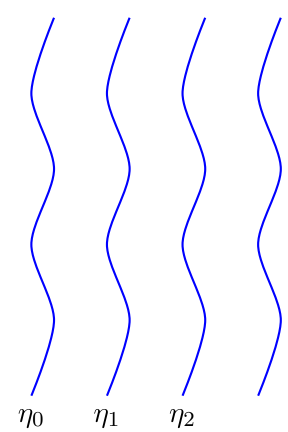

Figure 3. Family of $\{\eta_k\}_{k\in\mathbb{Z}}$

Clearly, these curves are mutually disjoint. These are similar to the ideas in Section 6.3.1, in which we used $\{\eta_k\}$ to create circle homeomorphisms. See Figure 3.

Step 3. Let ξ be a given orbit on $\mathcal{U}$. Then, in light of Lemma 7.9,

$$\lim_{t\to\infty} \frac{\xi(t)}{|\xi(t)|} = \lim_{t\to-\infty} \frac{-\xi(t)}{|\xi(t)|} = \vec{n}. \tag{7.8}$$

In particular, we yield that ξ intersects each η_k for $k \in \mathbb{Z}$. Unlike the smooth cases that ξ intersects η_k only once, the situation here is much more complicated. As a matter of fact, ξ might intersect η_k infinitely many times, and there might be pathological behaviors. It is thus extremely important to overcome this technical hurdle and to come up with a systematic way to understand the asymptotic behavior of ξ using $\{\eta_k\}$.

For $k \in \mathbb{Z}$, write

$$\begin{cases} t_{k,+} = \max\{t \in \mathbb{R} \,:\, \xi(t) \in \eta_k(\mathbb{R})\}, \\ t_{k,-} = \min\{t \in \mathbb{R} \,:\, \xi(t) \in \eta_k(\mathbb{R})\}. \end{cases}$$

Thanks to (7.8), both $t_{k,+}$ and $t_{k,-}$ are finite. See Figure 4.

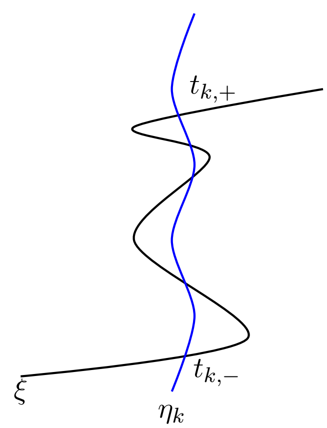

Figure 4. Intersections of η_k and ξ

For $k \in \mathbb{Z}$, assume that

$$\xi(t_{k,+}) = \eta_k(\theta_+) \quad \text{ and } \quad \xi(t_{k,-}) = \eta_k(\theta_-)$$

for $\theta_-, \theta_+ \in \mathbb{R}$. We claim that

$$\begin{cases} \xi(\mathbb{R}) \cap \eta_k(\mathbb{R}) \subset \{\eta_k(t) \,:\, \min\{\theta_+, \theta_-\} \le t \le \max\{\theta_+, \theta_-\}\}, \\ |\theta_+ - \theta_-| < T. \end{cases}$$

See Lemma 7.16 below for the proof.

Next, denote by

$$L_A = \int_0^T \left(\frac{1}{2} |\dot{\eta}(s)|^2 - V(\eta(s)) + c \right) ds,$$

which is exactly the action of one cycle of η. Pick

$$J > \max\left\{ \Lambda + 1,\ \frac{L_A}{\sqrt{c - \max_{\mathbb{T}^2} V}} + \Lambda \right\}.$$

Then,

$$t_{k,+} < t_{k+1,-} \quad \text{for all } k \in \mathbb{Z}.$$

See Lemma 7.17.

Step 4. For two orbits ξ_1 and ξ_2 on $\mathcal{U}$, we denote

$$d(\xi_1, \xi_2, k) = \min\{|\theta_1 - \theta_2| : \theta_1 \in A_{1,k},\ \theta_2 \in A_{2,k}\},$$

where, for $i = 1, 2$,

$$A_{i,k} = \{\theta \in \mathbb{R} : \eta_k(\theta) \in \xi_i(\mathbb{R})\}.$$

Then, define the distance

$$d(\xi_1, \xi_2) = \sum_{k \in \mathbb{Z}} \frac{\arctan(d(\xi_1, \xi_2, k))}{|k|^2 + 1}.$$

It is not hard to see that the distance functions $d(\cdot,\cdot,k)$ and $d(\cdot,\cdot)$ are lower semicontinuous with respect to orbits on $\mathcal{U}$, that is, if $\xi_{i,n} \to \xi_i$ locally uniformly for $i = 1, 2$, then

$$\liminf_{m \to \infty} d(\xi_{1,n}, \xi_{2,n}) \geq d(\xi_1, \xi_2).$$

Denote by

$$\mathcal{I} = \{\theta \in \mathbb{R} : \text{there exists an orbit } \xi \in \mathcal{U} \text{ such that } \eta(\theta) \in \xi(\mathbb{R})\}. \tag{7.9}$$

Due to the T-periodicity of η and the fact that the set $\mathcal{U}$ is closed under limits of orbits, $\mathcal{I}$ is a T-periodic closed subset of $\mathbb{R}$.

Step 5. We claim that

$$u_0 - u_1 \text{ is constant on } \mathcal{I}, \tag{7.10}$$

which gives to a contradiction as

$$\lim_{\theta \to \infty} |u_0(\eta(\theta)) - u_1(\eta(\theta))| = \infty.$$

Indeed, write

$$\mathbb{R} \backslash \mathcal{I} = \bigcup_{i=1}^{\infty} (a_i, b_i),$$

where $\{(a_i, b_i)\}_{i \geq 1}$ are disjoint open intervals. Obviously, $(a_i, b_i) \subseteq (a_i, a_i + T)$ for each $i \in \mathbb{N}$ since $\xi \in \mathcal{U} \Rightarrow \xi + (0,1) \in \mathcal{U}$. To prove (7.10), we assume first that, for $j \in \mathbb{N}$,

$$u_0(\eta(a_j)) - u_1(\eta(a_j)) = u_0(\eta(b_j)) - u_1(\eta(b_j)).$$

This will be verified in Lemma 7.18. Set, for $t \in \mathbb{R}$,

$$g(t) = u_0(\eta(t)) - u_1(\eta(t)).$$

Then, g is Lipschitz continuous and by (7.6),

$$g'(t) = 0 \quad \text{for } t \in \mathcal{I}.$$

Combine this with the fact that $g(a_i) = g(b_i)$ for $i \in \mathbb{N}$ to yield that

$$g(t) = c \quad \text{for } t \in \mathcal{I}$$

for some constant $c \in \mathbb{R}$. This is absurd as for $m \to \infty$,

$$|g(mT) - g(0)| \geq m|(p_0 - p_1) \cdot (0,1)| - \|v_0\|_{L^\infty(\mathbb{T}^2)} - \|v_1\|_{L^\infty(\mathbb{T}^2)} \to +\infty.$$

□

In the following, we present various results that were needed in the above proof. We are always in the setting of Theorem 7.15.

Lemma 7.16. *For $k \in \mathbb{Z}$, assume that*

$$\xi(t_{k,+}) = \eta_k(\theta_+) \qquad \textit{and} \qquad \xi(t_{k,-}) = \eta_k(\theta_-)$$

for $\theta_-, \theta_+ \in \mathbb{R}$. Then,

(1) $\xi(\mathbb{R}) \cap \eta_k(\mathbb{R}) \subset \{\eta_k(t) \,:\, \min\{\theta_+, \theta_-\} \leq t \leq \max\{\theta_+, \theta_-\}\}$;
(2) $|\theta_+ - \theta_-| < T$.

Proof. We first prove (1). Were the conclusion of (1) not true, there would exist $t_0 \in (t_{k,-}, t_{k,+})$ such that

$$\xi(t_0) = \eta_k(\theta)$$

for some $\theta < \min\{\theta_+, \theta_-\}$ or $\theta > \max\{\theta_+, \theta_-\}$. Without loss of generality, we assume that $\theta > \theta_+ > \theta_-$. Then,

$$\underbrace{\int_{\theta_-}^{\theta_+} \left(\frac{1}{2}|\dot{\eta}_k(s)|^2 - V(\eta_k(s)) + c\right) ds}_{A} = \underbrace{\int_{t_{k,-}}^{t_{k,+}} \left(\frac{1}{2}|\dot{\xi}(s)|^2 - V(\xi(s)) + c\right) ds}_{B}$$

as both η_k and ξ are absolute minimizers of the action connecting $\eta_k(\theta_-)$ and $\eta_k(\theta_+)$. On the other hand,

$$\underbrace{\int_{\theta_-}^{\theta} \left(\frac{1}{2}|\dot{\eta}_k(s)|^2 - V(\eta_k(s)) + c\right) ds}_{C} = \underbrace{\int_{t_{k,-}}^{t_0} \left(\frac{1}{2}|\dot{\xi}(s)|^2 - V(\xi(s)) + c\right) ds}_{D}$$

since both η_k and ξ are absolute minimizers of the action connecting $\eta_k(\theta_-)$ and $\eta_k(\theta)$. However, it is obvious that

$$B > D \qquad \text{and} \qquad A < C,$$

which gives a contradiction.

Next we prove (2). Again we argue by contradiction. Without loss of generality, assume that $\theta_+ - \theta_- \geq T$. See Figure 5.

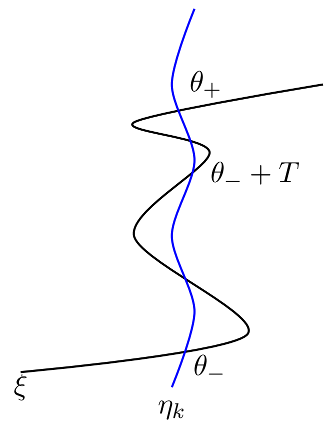

Figure 5. The situation where $\theta_+ - \theta_- \geq T$

Let ξ_1 be the adjustment of ξ with respect to η_k between $t_{k,-}$ and $t_{k,+}$ (see Definition 7.8). By (1),

$$\xi_1(\mathbb{R}) \cap \eta_k(\mathbb{R}) = \eta_k([\theta_-, \theta_+]).$$

In particular, ξ_1 contains two points $A = \eta_k(\theta_-)$ and $B = \eta_k(\theta_- + T) = A + (0,1)$. Thanks to Corollary 7.11, $\vec{n}$, the normal vector of F_c at p_0, is parallel to $(0,1)$, which contradicts the assumption that $\vec{n}$ is irrational. $\square$

Lemma 7.17. *Assume that*

$$J > \max\left\{\Lambda + 1,\ \frac{L_A}{\sqrt{c - \max_{\mathbb{T}^2} V}} + \Lambda\right\}.$$

Then,

$$t_{k,+} < t_{k+1,-} \qquad \textit{for all } k \in \mathbb{Z}.$$

Here, $\Lambda = \max\{|e_1 \cdot (x-y)| \,:\, x, y \in \eta(\mathbb{R})\}$.

Proof. Assume by contradiction that for some $k \in \mathbb{Z}$

$$t_{k+1,-} \in (t_{k,-}, t_{k,+}).$$

See Figure 6 for an illustration of this situation. Note first that

$$2(J - \Lambda) \leq |\xi(t_{k+1,-}) - \xi(t_{k,-})| + |\xi(t_{k+1,-}) - \xi(t_{k,+})| \leq \int_{t_{k,-}}^{t_{k,+}} |\dot{\xi}(s)|\, ds.$$

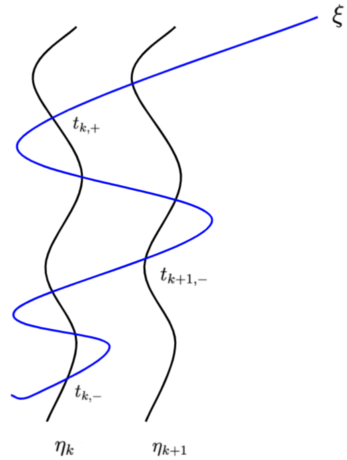

Figure 6. The situation where $t_{k+1,-} \in (t_{k,-}, t_{k,+})$

On the other hand,

$$\int_{t_{k,-}}^{t_{k,+}} \left(\frac{1}{2}|\dot{\xi}(s)|^2 - V(\xi(s)) + c\right) ds \geq \sqrt{2} \int_{t_{k,-}}^{t_{k,+}} \sqrt{c - V(\xi(s))}|\dot{\xi}(s)|\, ds$$
$$\geq \sqrt{c - \max_{\mathbb{T}^2} V} \int_{t_{k,-}}^{t_{k,+}} |\dot{\xi}(s)|\, ds.$$

Besides, for $\xi(t_{k,+}) = \eta_k(\theta_{k,+})$ and $\xi(t_{k,-}) = \eta_k(\theta_{k,-})$, we use (2) in Lemma 7.16 to yield that $|\theta_{k,+} - \theta_{k,-}| \leq T$. Hence,

$$\int_{t_{k,-}}^{t_{k,+}} \left(\frac{1}{2}|\dot{\xi}(s)|^2 - V(\xi(s)) + c\right) ds$$
$$= \left|\int_{\theta_{k,-}}^{\theta_{k,+}} \left(\frac{1}{2}|\dot{\eta}_k(s)|^2 - V(\eta_k(s)) + c\right) ds\right| \leq L_A.$$

Therefore,

$$2(J - \Lambda) \leq \frac{L_A}{\sqrt{c - \max_{\mathbb{T}^2} V}},$$

which contradicts the choice of J. □

Lemma 7.18. *For all $j \in \mathbb{N}$,*

$$u_0(\eta(a_j)) - u_1(\eta(a_j)) = u_0(\eta(b_j)) - u_1(\eta(b_j)). \tag{7.11}$$

Proof. We only need to prove the result for $j = 1$. Thanks to the lower semicontinuity of the distance function, we may choose ξ_1 and ξ_2 in $\mathcal{U}$ such that

$$\begin{cases} \xi_1(0) = \eta(a_1) \text{ and } \xi_2(0) = \eta(b_1), \\ d(\xi_1, \xi_2) \text{ is the smallest among all pair of curves satisfying the above.} \end{cases}$$

To simplify the associated topology between curves, we use Definition 7.8 to adjust ξ_1 and ξ_2 with respect to each η_k between $t_{i,k,-}$ and $t_{i,k,+}$ for $i = 1, 2$ respectively. Here

$$\begin{cases} t_{i,k,+} = \max\{t \in \mathbb{R} \,:\, \xi_i(t) \in \eta_k(\mathbb{R})\}, \\ t_{i,k,-} = \min\{t \in \mathbb{R} \,:\, \xi_i(t) \in \eta_k(\mathbb{R})\}. \end{cases}$$

By Lemma 7.17, for $i = 1, 2$, the time intervals $(t_{i,k,-}, t_{i,k,+})$ are mutually disjoint, that is,

$$... < t_{i,-1,-} \le t_{i,-1,+} < t_{i,0,-} \le t_{i,0,+} < t_{i,1,-} \le t_{i,1,+} < t_{i,2,-} \le t_{i,2,+} < ..$$

Hence the adjustments are well-defined.

In addition, thanks to (1) of Lemma 7.16, the distance between two adjusted orbits is not greater than $d(\xi_1, \xi_2)$. Thus two adjusted orbits also satisfy the above properties. By abuse of notations, we still use ξ_1 and ξ_2 to represent corresponding adjusted orbits, which mean

$$\xi_i(\mathbb{R}) \cap \eta_k(\mathbb{R}) = \xi_i([t_{i,k,-},\ t_{i,k,+}]) = \eta_k([\theta_{i,k,-},\ \theta_{i,k,+}]). \tag{7.12}$$

Here $\xi_i(t_{i,k,+}) = \eta_k(\theta_{i,k,+})$ and $\xi(t_{i,k,-}) = \eta_k(\theta_{i,k,-})$. It could happen that $\theta_{i,k,+} < \theta_{i,k,-}$. In terms of topology, the above adjustment basically plays the role like that ξ_i and η_k only intersect once for smooth V. This helps to avoid pathological behaviors about the intersections between ξ_i and η_k. We consider two cases.

Case 1. $\xi_1(\mathbb{R}) \cap \xi_2(\mathbb{R}) \ne \emptyset$, that is, ξ_1 and ξ_2 intersect. Assume that $\xi_1(t_1) = \xi_2(t_2)$ for some $t_1, t_2 \in \mathbb{R}$. See Figure 7. Then,

$$\begin{aligned} u_0(\xi_1(t_1)) - u_0(\xi_1(0)) &= u_1(\xi_1(t_1)) - u_1(\xi_1(0)) \\ &= \int_0^{t_1} \left(\frac{1}{2}|\dot\xi_1(s)|^2 - V(\xi_1(s)) + c\right) ds, \end{aligned}$$

and

$$\begin{aligned} u_0(\xi_2(t_2)) - u_0(\xi_2(0)) &= u_1(\xi_2(t_2)) - u_1(\xi_2(0)) \\ &= \int_0^{t_2} \left(\frac{1}{2}|\dot\xi_2(s)|^2 - V(\xi_2(s)) + c\right) ds. \end{aligned}$$

Taking the difference of the two equalities above leads to the desired result.

Case 2. $\xi_1(\mathbb{R}) \cap \xi_2(\mathbb{R}) = \emptyset$, that is, ξ_1 and ξ_2 do not intersect. For each $k \in \mathbb{Z}$, let

$$d_1(k) = \max\{\theta \,:\, \eta_k(\theta) \in \xi_1(\mathbb{R}) \cap \eta_k(\mathbb{R})\},$$

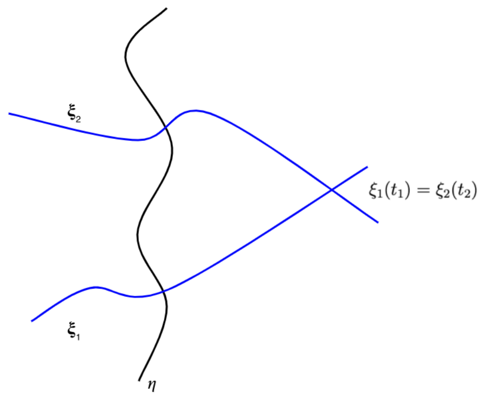

Figure 7. The situation where $\xi_1(\mathbb{R}) \cap \xi_2(\mathbb{R}) \neq \emptyset$

and

$$d_2(k) = \min\{\theta \,:\, \eta_k(\theta) \in \xi_2(\mathbb{R}) \cap \eta_k(\mathbb{R})\}.$$

In lights of the two dimensional topology and (7.12), we have that

$$d_1(0) = a_1 \qquad \text{and} \qquad d_2(0) = b_1,$$

and

$$d_1(k) < d_2(k) \qquad \text{for all } k \in \mathbb{Z}.$$

See Figure 8. We claim that

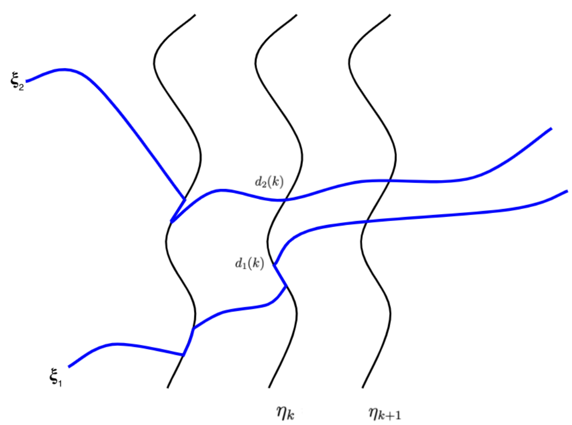

Figure 8. Positions of $d_1(k), d_2(k)$

$$\sum_{k\in\mathbb{Z}}(d_2(k)-d_1(k)) \le T. \tag{7.13}$$

Indeed, to verify this claim, we first show that, for all $k \in \mathbb{Z}$, the open interval

$$(d_1(k),\ d_2(k)) \subset \mathbb{R}\backslash\mathcal{I}, \tag{7.14}$$

that is, it is one of those open intervals $\{(a_j, b_j)\}_{j\ge1}$ described earlier. It suffices to show this for $k > 0$ as the proof for $k < 0$ is similar. We argue by contradiction. If this were not true, then there would exist $k \in \mathbb{N}$ and $\tilde{\xi} \in \mathcal{U}$ such that

$$\tilde{\xi}(0) \in \{\eta_k(\theta) : \theta \in (d_1(k),\ d_2(k))\}. \tag{7.15}$$

Since $\tilde{\xi}$ cannot pass the portion of η on (a_0, b_0), we deduce that, if trace backward along $\tilde{\xi}$, it must intersect ξ_1 or ξ_2 before it intersects η. See Figure 9.

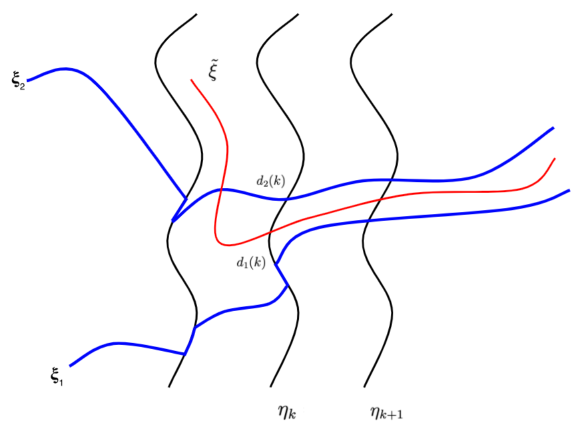

Figure 9. Relative position of $\tilde{\xi}$

Let

$$t_- = \max\{t \le 0 : \tilde{\xi}(t) \in \xi_1(\mathbb{R}) \cup \xi_2(\mathbb{R})\},$$
$$t_+ = \inf\{t \ge 0 : \tilde{\xi}(t) \in \xi_1(\mathbb{R}) \cup \xi_2(\mathbb{R})\}.$$

Then $t_- < 0$ and $t_+ > 0$. Note that t_+ could be $+\infty$.

Now we will use the gluing property of Definition 7.7 to construct a new orbit in $\mathcal{U}$. By two dimensional topology and (7.12), it is easy to see that, for each $k \in \mathbb{Z}$,

$$\tilde{\xi}((t_-, t_+)) \cap (\eta_k(\mathbb{R})) \subset \{\eta_k(\theta) : \theta \in (d_1(k),\ d_2(k))\}. \tag{7.16}$$

Without loss of generality, we assume that $\tilde{\xi}(t_-) \in \xi_2(\mathbb{R})$ and $\tilde{\xi}(t_+) \in \xi_j(\mathbb{R})$ for $j = 1$ or $j = 2$ if $t_+ < +\infty$. Suppose that

$$\tilde{\xi}(t_-) = \xi_2(\bar{t}_-) \qquad \text{and} \qquad \tilde{\xi}(t_+) = \xi_j(\bar{t}_+).$$

for $0 < \bar{t}_- < \bar{t}_+$. If $t_+ = +\infty$, then denote by $\bar{t}_+ = +\infty$. Let

$$\xi_3(t) = \begin{cases} \xi_2(t) & \text{for } t \le \bar{t}_-, \\ \tilde{\xi}(t + t_- - \bar{t}_-) & \text{for } \bar{t}_- \le t \le t_+ + \bar{t}_- - t_-, \\ \xi_j(t + \bar{t}_+ + t_- - \bar{t}_- - t_+) & \text{for } t \ge t_+ + \bar{t}_- - t_-. \end{cases}$$

We then use (7.15) and (7.16) to deduce that

$$\xi_3(0) = \eta(b_1) \qquad \text{and} \qquad d(\xi_3, \xi_1) < d(\xi_2, \xi_1),$$

which contradicts the choice of ξ_1 and ξ_2. Hence our claim (7.14) holds.

Next we show that for $k \ne l$, $(d_1(k), d_2(k))$ is not a T-translation of $(d_1(l), d_2(l))$. In fact, if

$$(d_1(k), d_2(k)) = (d_1(l), d_2(l)) + jT$$

for some $j \in \mathbb{Z}\backslash\{0\}$, then both $\eta(d_1(k))$ and $\eta(d_1(l)) + ((k-l)J, j)$ are on ξ_1. Then, the outward normal vector $\vec{n}$ is rational, which contradicts our assumption. Accordingly, after we translate all $(d_1(k), d_2(k))$ into $(a_1, a_1 + T)$, they are all disjoint. Therefore, (7.13) holds true.

The property (7.13) implies

$$\lim_{k\to\infty} (d_2(k) - d_1(k)) = 0.$$

Similar to Case 1 above, since $\eta_k(d_i(k)) \in \xi_i(\mathbb{R})$ for $i = 1, 2$,

$$u_0(\eta_k(d_1(k))) - u_0(\xi_1(0)) = u_1(\eta_k(d_1(k))) - u_1(\xi_1(0)),$$

and

$$u_0(\eta_k(d_2(k))) - u_0(\xi_2(0)) = u_1(\eta_k(d_2(k))) - u_1(\xi_2(0)).$$

Taking the difference of the two equations and sending $k \to \infty$, we obtain the claim.

$\square$

7.3. Effective fronts in two dimensions

We are always in the two dimensional setting, that is, $n = 2$ in this section. In this section, we focus on the front propagation Hamiltonian

$$H(y, p) = a(y)|p| \qquad \text{for } (y, p) \in \mathbb{T}^2 \times \mathbb{R}^2$$

for some $a \in C(\mathbb{T}^2, (0, \infty))$. Again, we do not require any smoothness of a here.

Denote by $\overline{H}_a = \overline{H}$ the corresponding effective Hamiltonian. We write $\overline{H}_a$ to demonstrate clearly the dependence on a. Of course, $\overline{H}_a$ is positive homogeneous of degree one, and its 1-sublevel set

$$S_a := \left\{ p \in \mathbb{R}^2 \, : \, \overline{H}_a(p) \leq 1 \right\}$$

belongs to $\mathcal{W}$, which denotes the collection of all convex sets in $\mathbb{R}^2$ that are centrally symmetric with nonempty interior. The convex dual D_a of S_a, determined by

$$D_a = \partial \overline{H}_a(0),$$

the subdifferential of $\overline{H}_a$ at the origin, is called the effective front, which also belongs to $\mathcal{W}$.

7.3.1. Properties of the effective front.

Theorem 7.19. *Assume that $n = 2$, and $H(y,p) = a(y)|p|$ for $(y,p) \in \mathbb{T}^2 \times \mathbb{R}^2$ for some $a \in C(\mathbb{T}^2, (0,\infty))$. Then, ∂S_a does not contain a line segment of irrational slope. Equivalently, ∂D_a is differentiable at every irrational point.*

Proof. We use Theorem 7.15 to obtain the result. We consider the closely related mechanical Hamiltonian

$$K(y,p) = \frac{1}{2}|p|^2 + V(y) \qquad \text{for } (y,p) \in \mathbb{T}^2 \times \mathbb{R}^2,$$

for some $V \in C(\mathbb{T}^2)$ to be chosen. Let $\overline{K}$ be the associated effective Hamiltonian. Of course, $\min \overline{K} = \max_{\mathbb{T}^2} V$. For any $p \in \mathbb{R}^2$ with $\overline{K}(p) > \max_{\mathbb{T}^2} V$,

$$\frac{1}{2}|p + Dv|^2 + V(y) = \overline{K}(p) \qquad \Longleftrightarrow \qquad \frac{1}{\sqrt{2(\overline{K}(p) - V(y))}}|p + Dv| = 1.$$

Pick $c = 0$, and

$$V(y) = -\frac{1}{2a(y)^2} \qquad \text{for } y \in \mathbb{T}^2.$$

Then, $c > \max_{\mathbb{T}^2} V$, and

$$a(y) = \frac{1}{\sqrt{2(-V(y))}} = \frac{1}{\sqrt{2(c - V(y))}}.$$

By the above relation,

$$F_0 = \left\{ p \in \mathbb{R}^2 \, : \, \overline{K}(p) = 0 \right\} = \partial S_a.$$

By Theorem 7.15, F_0 does not contain a line segment of irrational slope.

$\square$

7.3.2. Constructions of polygonal effective fronts with rational vertices. We next have the following result.

Theorem 7.20. *Assume that $n = 2$, and $H(y,p) = a(y)|p|$ for $(y,p) \in \mathbb{T}^2 \times \mathbb{R}^2$ for some $a \in C(\mathbb{T}^2,(0,\infty))$. Then, for any $\alpha \in (0,1)$ and for any centrally symmetric polygon P with rational slopes and nonempty interior, there exists $a \in C^{1,\alpha}(\mathbb{T}^2,(0,\infty))$ such that*

$$S_a = P.$$

Note that it is not possible to have S_a of polygonal shape if a is C^2. We already proved in Theorem 6.10 that if a is C^2 and not constant, then S_a is C^1 and contains some flat pieces. Therefore, the result in Theorem 7.20 is optimal in terms of both the regularity of a and the obtainable shapes of S_a.

We proceed to prove Theorem 7.20 in this subsection. We need to have various preparation steps first.

Let P be a given centrally symmetric polygon with rational slopes $\{q_i\}_{i=1}^m$. As we are in two dimensions, we assume that the rational vectors $\{q_i\}_{i=1}^m \subset \mathbb{R}^2$ are arranged clockwise as in Figure 10. For each $i = 1,..,m$, there are a unique $\lambda_i > 0$ and a unique irreducible integer vector $(m_i,n_i) \in \mathbb{Z}^2$ so that

$$q_i = \lambda_i\,(m_i,n_i).$$

Of course, $\{q_i\}_{i=1}^m$ form normal vectors of half of the edges of P. We order

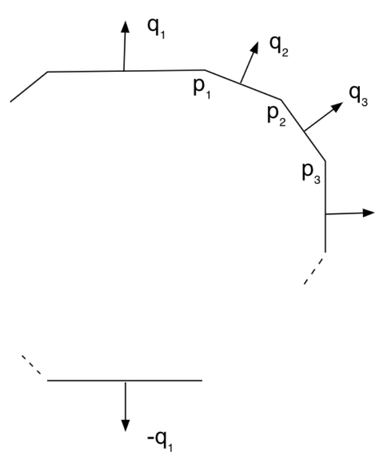

Figure 10. Polygon P with vertices $p_1, p_2, \ldots, p_{2m}$

the other half by

$$q_{m+i} = -q_i, \qquad 1 \le i \le m.$$

Let p_i be the vertex between q_i and q_{i+1} for $1 \le i \le 2m-1$. Let p_{2m} be the vertex between q_{2m} and q_1. We then have the following relations by normalizing $\{q_i\}_{i=1}^m$ appropriately, for $1 \le i \le m$,

$$p_i \cdot q_i = p_i \cdot q_{i+1} = 1 \qquad \text{and} \qquad \max_{\substack{j \neq i, i+1 \\ 1 \le j \le m}} |q_j \cdot p_i| < 1 = p_i \cdot q_i. \tag{7.17}$$

Lemma 7.21. *Suppose that $\xi \in C^1([0,T], \mathbb{R}^2)$ satisfies that*

$$\xi(T) - \xi(0) = (m,n) \in \mathbb{Z}^2.$$

Denote by

$$\lambda = \left(\int_0^T \frac{|\dot{\xi}(t)|}{a(\xi(t))}\, dt \right)^{-1}.$$

Then

$$\overline{H}_a(p) \ge \lambda\, p \cdot (m,n).$$

Proof. Thanks to the inf-max formula, it suffices to show that for any $\phi \in C^\infty(\mathbb{T}^2)$,

$$M := \max_{x \in \mathbb{R}^2} a(x)|p + D\phi(x)| \ge \lambda\, p \cdot (m,n).$$

Let $u(x) = p \cdot x + \phi(x)$ for $x \in \mathbb{R}^2$. We see that

$$p \cdot (m,n) = u(\xi(T)) - u(\xi(0)) = \int_0^T Du(\xi(t)) \cdot \dot{\xi}(t)\, dt \le \frac{M}{\lambda},$$

which yields the needed inequality. $\square$

We now create a suitable network with directions $\{q_i\}_{i=1}^m$. Choose m lines $\{L_i\}_{i=1}^m$ in $\mathbb{R}^2$ such that L_i is parallel to q_i for $1 \le i \le m$, and, when projected to $\mathbb{T}^2$, no three lines intersect at one point. Note that the projection of each line to $\mathbb{T}^2$ gives a periodic orbit. By (7.17), for every two distinct points x and y on L_i, we have that

$$|p_i \cdot (x-y)| > \max_{\substack{j \neq i-1, i \\ 1 \le j \le m}} |p_j \cdot (x-y)|.$$

Consider all integer translations of L_i, which form a network

$$\Upsilon = \bigcup_{i=1}^m \left(L_i + \mathbb{Z}^2\right).$$

Let I be the collection of all intersection points in this network Υ. Of course, I is $\mathbb{Z}^2$-periodic. Denote by

$$d = \min\{|x-y| \, : \, x \neq y, \ x, y \in I\}.$$

Clearly, $d > 0$.

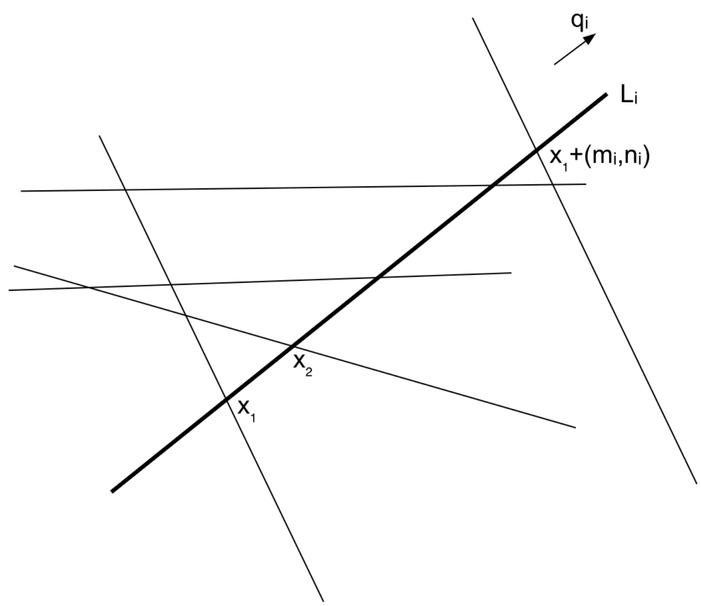

Figure 11. Intersection points on L_i

Next, we perform the following procedure at each point in I. In a small neighborhood of each fixed intersection point in I, we perturb the two corresponding intersecting lines a bit to create gradient flows of an appropriate function. As this point is of distance at least d away from other intersection points, this process is purely local. By linear transformations and translations, it suffices to show how to perform this procedure in a neighborhood of the origin $(0,0)$ provided that L_1, L_2 are the x_1-axis and x_2-axis, respectively. The adjustment can be done by using the following lemma.

Lemma 7.22. *Let $\alpha \in (0,1)$ be a fixed number as in the statement of Theorem 7.20. Pick $k \in \mathbb{N}$ so that*

$$\alpha \leq 1 - \frac{1}{2k}.$$

Consider the potential function

$$u(x_1,x_2) = C_k \left(\frac{x_1^{4k}}{C_k} + x_2^2 \right)^{1-\frac{1}{4k}} + 2x_1 \qquad \text{for } (x_1,x_2) \in \mathbb{R}^2,$$

where $C_k > 2k(4k+1)$ is a constant. Then, u has infinitely many distinct gradient flows passing through the origin.

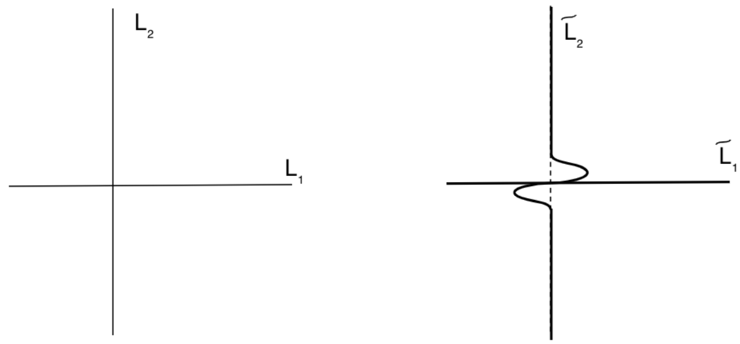

Figure 12. Local perturbation at the intersection of L_1 and L_2

Proof. Clearly, $u \in C^{1,1-\frac{1}{2k}}(\mathbb{R}^2)$ and is C^2 away from the origin.

Firstly, we note that $\gamma_1(t) = (f(t), 0)$ with

$$\begin{cases} f'(t) = 2 + C_k^{\frac{1}{4k}}(4k-1)f(t)^{4k-2}, \\ f(0) = 0. \end{cases}$$

is a gradient flow of u passing through the origin.

Denote by

$$D = \left\{(a, b) \,:\, 0 < a < 1,\ 0 < b < a^{2k}\right\}.$$

To finish, it suffices to show that if $\xi(t) = (x_1(t), x_2(t)) : \mathbb{R} \to \mathbb{R}^2$ is a gradient flow of u and $\xi(0) \in D$, then

$$\xi((-\infty, 0)) \cap (0, \infty)^2 \subset D.$$

Note that $x_1(t)$ and $x_2(t)$ are both increasing within D and ξ cannot intersect with γ_1 away from the origin. If the above statement were not correct, there would exist $\theta < 0$ such that

$$0 < x_2(\theta) = x_1^{2k}(\theta) \quad \text{and} \quad 0 < x_2(t) < x_1^{2k}(t) < 1 \quad \text{for } t \in (\theta, 0).$$

At θ,

$$\frac{C_k x_1^{2k-1}(\theta)}{1+4k} < \frac{u_{x_2}(x_1(\theta), x_2(\theta))}{u_{x_1}(x_1(\theta), x_2(\theta))} = \frac{x_2'(\theta)}{x_1'(\theta)} \leq 2k x_1^{2k-1}(\theta),$$

which contradicts the assumption that $C_k > 2k(4k+1)$. The proof is complete. $\square$

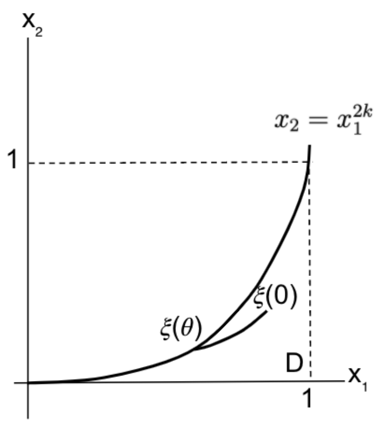

Figure 13. Graph of ξ in D

By the above constructions, we get m periodic curves $\{\tilde{L}_i\}_{i=1}^m$ and their integer translations such that, for some small $r \in (0, \frac{d}{10})$, we have the following important properties.

(1) $\tilde{L}_i = L_i$ away from the set $I_r = \{x \in \mathbb{R}^2 : d(x, I) \le r\}$.

(2) The set of intersection points remains the same, that is, for $i \neq j$ and any integer vector $v \in \mathbb{Z}^2$,

$$\tilde{L}_i \cap (\tilde{L}_j + v) = L_i \cap (L_j + v).$$

Equivalently, $\tilde{L}_i \cap \tilde{L}_j = L_i \cap L_j$ when projected to $\mathbb{T}^2$.

(3) For $i \neq j$ and an integer vector $v \in \mathbb{Z}^2$, if $\tilde{L}_i$ and $\tilde{L}_j + v$ intersect at $x = x_{i,j,v}$, then there exists a $C^{1,\alpha}$ function $u = u_{i,j,v}$ in $B_{\frac{r}{2}}(x)$ such that

- $|Du(x)| \ge 1$ in $B_{\frac{r}{2}}(x)$;
- within $B_{\frac{r}{2}}(x)$, $\tilde{L}_i$ and $\tilde{L}_j + v$ are two gradient flows of u that only intersect at x;
- (periodicity) if two intersection points $x_{i,j,v} = x_{i',j',v'} + w$ for some $w \in \mathbb{Z}^2$, then

$$u_{i,j,v}(x + w) = u_{i',j',v'}(x) \quad \text{for } x \in B_r(x_{i',j',v'}).$$

In particular, u is well defined on $I_{\frac{r}{2}}$ when being projected to the flat torus $\mathbb{T}^2$.

Denote by

$$\Gamma = \bigcup_{1\le i\le m} (\tilde{L}_i + \mathbb{Z}^2)$$

the perturbed network.

We are done with important preparation results. We now give the proof of the main theorem in this subsection.

Proof of Theorem 7.20. As usual, we divide the proof into several steps.

Step 1. Initial choice of a_0. We first pick $r_0 \in (0, \frac{r}{2})$ and $a_0 \in C^{1,\alpha}(\mathbb{T}^2, (0,\infty))$ such that a_0 is C^∞ away from the set I and satisfies the following conditions.

(1) For every intersection point $x = x_{i,j,v} \in I$ and the associated function $u = u_{i,j,v}$ from the above construction, let

$$a_0(y) = \frac{1}{|Du(y)|} \qquad \text{for } x \in B_{r_0}(x).$$

(2) For every two intersection points x, y on $\tilde{L}_i$ for $1 \le i \le m$ (i.e., $x, y \in \tilde{L}_i \cap I$), the weighted length $l_i(x, y)$ between x and y along $\tilde{L}_i$ satisfies

$$l_i(x,y) := \int_0^1 \frac{1}{a_0(\xi(t))} |\dot{\xi}(t)|\, dt = |p_i \cdot (x - y)|. \tag{7.18}$$

Here, $\xi : [0,1] \to \tilde{L}_i$ is an arbitrary parametrization of $\tilde{L}_i$ between x and y. In particular, the weighted length of each period (i.e., from x to $x+(m_i, n_i)$) of $\tilde{L}_i$ is

$$|p_i \cdot (m_i, n_i)| = \frac{1}{\lambda_i}|p_i \cdot q_i| = \frac{1}{\lambda_i}.$$

The existence of a_0 is clear provided $r > 0$ is small enough. By Lemma 7.21,

$$\overline{H}_{a_0}(p) \ge \max_{1\le i\le m} |q_i \cdot p|. \tag{7.19}$$

For $i = 1, 2, ..., m$, let $\xi_i : \mathbb{R} \to \tilde{L}_i$ be the smooth reparametrization of $\tilde{L}_i$ such that

$$|\dot{\xi}_i(t)| = \frac{1}{a_0(\xi_i(t))} \qquad \text{for } t \in \mathbb{R}.$$

By usual constructions and the periodicity of Γ, there exists a universal $\delta_0 \in (0, r_0)$ such that for each $i = 1, 2, 3, ..., m$, there exists $w_i \in C^{1,\alpha}(\tilde{L}_{i,\delta_0})$ such that w_i is C^∞ away from intersection points and

(1) $\dot{\xi}_i(t) = Dw_i(\xi_i(t))$ for all $t \in \mathbb{R}$, i.e., ξ_i is the gradient flow of w_i;

(2) $Dw_i(x) = Du_{i,j,v}(x)$ for $x \in B_{\delta_0}(x_{i,j,v})$;

(3) $\inf_{x\in \tilde{L}_{i,\delta_0}} |Dw_i(x)| > 0$.

Here, $\tilde{L}_{i,\delta_0} = \{x : d(x, \tilde{L}_i) < \delta_0\}$ and $x_{i,j,v}$ is any intersection point on $\tilde{L}_i$. Let

$$\Gamma_{\delta_0} = \{x \in \mathbb{R}^2 : d(x, \Gamma) < \delta_0\} = \bigcup_{i=1}^{m} (\tilde{L}_{i,\delta_0} + \mathbb{Z}^2).$$

Then, for $x \in \Gamma_{\delta_0}$, we define

$$a_0(x) = \frac{1}{|Dw_i(x - v)|} \qquad \text{if } x - v \in \tilde{L}_{i,\delta_0} \text{ for } 1 \le i \le m, \text{ and } v \in \mathbb{Z}^2.$$

Extend a_0 to $C^{1,\alpha}(\mathbb{T}^2, (0, \infty))$ in such a way that it is smooth away from I.

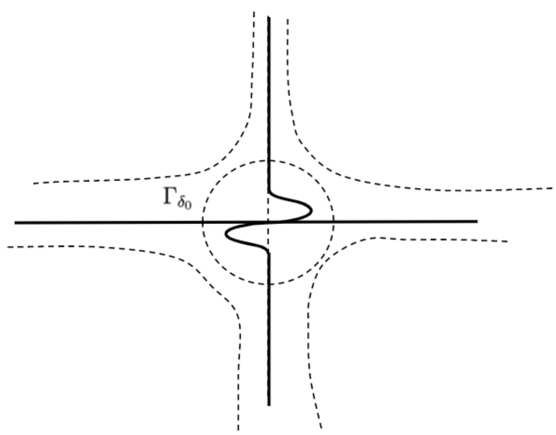

Figure 14. Part of Γ_{δ_0}

Step 2. Adjustments of a_0. Next we need to construct $\tilde{a} \in C^{1,\alpha}(\mathbb{T}^2, (0, \infty))$ that is smooth away from I and satisfies

$$\begin{cases} \tilde{a} = a_0 & \text{on } \Gamma, \\ \overline{H}_{\tilde{a}}(p_i) \le 1 & \text{for } 1 \le i \le m. \end{cases}$$

Since $\tilde{a}$ agree with a_0 of the previous step along $\tilde{L}_i$'s, the property (7.18) and, by Lemma 7.21, the inequality (7.19) are preserved. We hence obtain

$$\begin{cases} \overline{H}_{\tilde{a}}(p) \ge \max_{1 \le i \le m} |q_i \cdot p|, \\ \overline{H}_{\tilde{a}}(p_i) \le 1 \qquad \text{for } 1 \le i \le m. \end{cases}$$

Therefore, the function $\tilde{a}$ is exactly what we are looking for, that is, $S_{\tilde{a}} = P$.

Let us now give the construction of $\tilde{a}$. In light of (7.17), for given $i \in \{1, 2, 3, .., m\}$, the following points hold

- for $j = i, i+1$ and two intersection points $x, y \in L_j$,

$$|p_i \cdot x - p_i \cdot y| = l_j(x, y);$$

- for $j \neq i, i+1$ and every two distinct intersection points $x, y \in L_j$,

$$|p_i \cdot x - p_i \cdot y| = |p_i \cdot (x-y)| \leq \max_{l \neq j-1, j} |p_l \cdot (x-y)| < |p_j \cdot (x-y)| = l_j(x, y).$$

By usual constructions and the periodicity of Γ, there exists $\mu_0 \in (0, \delta_0)$ such that for each $i = 1, 2, 3, .., m$, there exists a function $\tilde{u}_i \in C^{1,\alpha}(\Gamma_{\mu_0})$ such that

$$\begin{cases} \tilde{u}_i \in C^{1,\alpha}(\Gamma_{\delta_0}), \ \tilde{u}_i \in C^{\infty}(\Gamma_{\mu_0} \backslash I), \\ \inf_{\Gamma_{\delta_0}} |D\tilde{u}_i| > 0, \\ \tilde{u}_i - p_i \cdot x \quad \text{is } \mathbb{Z}^2\text{-periodic in } \Gamma_{\mu_0}, \\ |D\tilde{u}_i| \leq |Dw_i| \quad \text{in } \Gamma_{\mu_0}, \end{cases}$$

and for any intersection point $x = x_{i,j,v} \in I$,

$$D\tilde{u}_i = Dw_i = Du_{i,j,v} \qquad \text{in } B_{\mu_0}(x_{i,j,v}).$$

We extend $\tilde{u}_i - p_i \cdot x$ to $v_i \in C^{1,\alpha}(\mathbb{T}^2)$ such that v_i is C^2 away from I, and for $u_i = p_i \cdot x + v_i$,

$$u_i = \tilde{u}_i \qquad \text{on } \Gamma_{\frac{\mu_0}{2}}.$$

Now let

$$K_1 = \max_{1 \leq i \leq m} \max_{x \in \mathbb{R}^2} |Du_i(x)| \qquad \text{and} \qquad K_2 = \max_{x \in \mathbb{R}^2} a_0(x).$$

Choose a cut-off function $\phi(x) \in C^{\infty}(\mathbb{T}^2, (0, 1])$ such that

$$\phi(x) = \begin{cases} 1 & \text{for } x \in \Gamma_{\frac{\mu_0}{4}}, \\ \frac{1}{K_1(1+K_2)} & \text{for } x \in \mathbb{R}^2 \backslash \Gamma_{\frac{\mu_0}{2}}. \end{cases}$$

We then simply define

$$\tilde{a}(x) = \phi(x) a_0(x) \qquad \text{for } x \in \mathbb{R}^2.$$

Then, for $i = 1, 2, ..., m$,

$$\begin{cases} \tilde{a}(x)|p + Dv_i(x)| \leq \tilde{a}(x)|Dw_i(x)| = \phi(x) \leq 1 & \text{for } x \in \Gamma_{\frac{\mu_0}{2}}, \\ \tilde{a}(x)|p + Dv_i(x)| = \frac{a_0(x)|Du_i(x)|}{K_1(1+K_2)} \leq 1 & \text{for } x \in \mathbb{R}^2 \backslash \Gamma_{\frac{\mu_0}{2}}, \end{cases}$$

which implies

$$\max_{x \in \mathbb{R}^2} \tilde{a}(x)|p + Dv_i(x)| = \max_{x \in \mathbb{R}^2} \tilde{a}(x)|Du_i(x)| \leq 1.$$

By the inf-max formula, for $1 \le i \le m$,

$$\overline{H}_{\tilde{a}}(p_i) \le 1.$$

Thus, $\tilde{a}$ constructed above has the desired properties. The proof of Theorem 7.20 is complete.

$\square$

Thanks to the two main theorems, Theorem 7.19 and Theorem 7.20, we have the following important claim.

Theorem 7.23. *Assume that $n = 2$, and $H(y,p) = a(y)|p|$ for $(y,p) \in \mathbb{T}^2 \times \mathbb{R}^2$ for some $a \in C(\mathbb{T}^2, (0,\infty))$. Then, a polygon could be an effective front D_a if and only if it is centrally symmetric with rational vertices and nonempty interior.*

7.4. Open problems

In the following, we list several open problems along the directions discussed in this chapter.

Question 1. *In Lemma 7.1 , we obtained that for all $t \in (a,b)$, $w_1 - w_2$ is differentiable at $x = \eta(t)$, and*

$$D(w_1 - w_2)(x) = 0.$$

Is it possible to show that w_1 and w_2 are individually differentiable at $x = \eta(t)$ for all $t \in (a,b)$?

Next, we address questions concerning the shape of $\overline{H}_a$ and the effective fronts. For clarity, recall that

$$S_a := \left\{p \in \mathbb{R}^2 \,:\, \overline{H}_a(p) \le 1\right\}$$

belongs to $\mathcal{W}$, the collection of all convex sets in $\mathbb{R}^2$ that are centrally symmetric with nonempty interior. The convex dual D_a of S_a, determined by

$$D_a = \partial \overline{H}_a(0),$$

the subdifferential of $\overline{H}_a$ at the origin, is called the effective front, which also belongs to $\mathcal{W}$.

Question 2. *Is any set in $\mathcal{W}$ realizable as the effective front if we look at $a \in L^\infty(\mathbb{T}^2, (0,\infty))$ with $\operatorname{ess\,inf}_{\mathbb{T}^2} a > 0$?*

Question 3. *Does there exist a nonconstant $a \in C(\mathbb{T}^2, (0,\infty))$ such that S_a is a strictly convex set (e.g., a disk)?*

Question 4. *Assume $n \ge 3$. What are the differentiability properties of $\overline{H}_a$? What can we say about the effective front?*

7.5. References

(1) The first part of this chapter is based on Tran, Yu [**TY22**].

(2) The second part of this chapter is based on Jing, Tran, Yu [**JTY21**].

Chapter 8

Optimal rate of convergence for periodic homogenization of Hamilton-Jacobi equations in the convex setting

In this chapter, we obtain optimal rate of convergence for periodic homogenization of Hamilton-Jacobi equations in the convex setting. We always assume that the Hamiltonian $H = H(y,p) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a given continuous function satisfying

$$(8.1) \qquad \begin{cases} y \mapsto H(y,p) \text{ is } \mathbb{Z}^n\text{-periodic for each } p \in \mathbb{R}^n, \\ \lim_{|p|\to\infty} \min_{y\in\mathbb{R}^n} H(y,p) = +\infty, \\ p \mapsto H(y,p) \text{ is convex for each } y \in \mathbb{T}^n. \end{cases}$$

Of course, a short way to state (8.1) is that $H \in C(\mathbb{T}^n \times \mathbb{R}^n)$ is convex and coercive in p. Let us now give a very quick summary of the qualitative theory of periodic homogenization of Hamilton-Jacobi equations. In fact, we do not need to use this qualitative theory here and we obtain directly the optimal quantitative estimates. Nevertheless, it is important to mention things clearly here.

For each $\varepsilon > 0$, let $u^\varepsilon \in C(\mathbb{R}^n \times [0,\infty))$ be the viscosity solution to

$$\begin{cases} u^\varepsilon_t + H\left(\frac{x}{\varepsilon}, Du^\varepsilon\right) = 0 & \text{in } \mathbb{R}^n \times (0,\infty), \\ u^\varepsilon(x,0) = g(x) & \text{on } \mathbb{R}^n. \end{cases} \tag{8.2}$$

For the initial data, we typically assume $g \in \mathrm{BUC}\,(\mathbb{R}^n) \cap \mathrm{Lip}\,(\mathbb{R}^n)$. As $\varepsilon \to 0$, u^ε converges to u locally uniformly on $\mathbb{R}^n \times [0,\infty)$ as $\varepsilon \to 0$, and u solves the effective equation

$$\begin{cases} u_t + \overline{H}\,(Du) = 0 & \text{in } \mathbb{R}^n \times (0,\infty), \\ u(x,0) = g(x) & \text{on } \mathbb{R}^n. \end{cases} \tag{8.3}$$

Here, $\overline{H}$ is the usual effective Hamiltonian.

Our main goal in this chapter is to obtain rate of convergence of u^ε to u in L^∞, that is, an optimal bound for $\|u^\varepsilon - u\|_{L^\infty(\mathbb{R}^n \times [0,\infty))}$ as $\varepsilon \to 0+$. Here is the main result.

Theorem 8.1. *Assume* (8.1) *and* $g \in \mathrm{BUC}\,(\mathbb{R}^n) \cap \mathrm{Lip}\,(\mathbb{R}^n)$. *For* $\varepsilon > 0$, *let* u^ε *be the viscosity solution to* (8.2). *Let* u *be the viscosity solution to* (8.3). *Then, there exists* $C > 0$ *depending only on* H, $\|Dg\|_{L^\infty(\mathbb{R}^n)}$, *and* n *such that*

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

For $(x,t) \in \mathbb{R}^n \times (0,\infty)$, the optimal control formula for the solution to (8.2) is

$$u^\varepsilon(x,t) = \inf_{\substack{\gamma(t)=x \\ \gamma \in \mathrm{AC}\,([0,t])}} \left\{ g\left(\gamma\left(0\right)\right) + \int_0^t L\left(\frac{\gamma(s)}{\varepsilon}, \dot\gamma(s)\right) ds \right\}.$$

It is hard to understand the action with the highly oscillatory spatial variable $\frac{\gamma(s)}{\varepsilon}$. We make a change of variables to see this averaging effect better. Denote by

$$r = \frac{s-t}{\varepsilon} \qquad \text{for } -\frac{t}{\varepsilon} \le r \le 0,$$

and

$$\eta(r) = \frac{\gamma(\varepsilon r)}{\varepsilon} \qquad \Longrightarrow \qquad \dot\eta(r) = \dot\gamma(\varepsilon r).$$

Then, the above optimal control formula becomes

$$u^\varepsilon(x,t) = \inf_{\substack{\varepsilon\eta(0)=x \\ \eta \in \mathrm{AC}\,([-\varepsilon^{-1}t,0])}} \left\{ g\left(\varepsilon\eta\left(-\varepsilon^{-1}t\right)\right) + \varepsilon \int_{-\varepsilon^{-1}t}^0 L(\eta(r),\dot\eta(r))\,dr \right\}. \tag{8.4}$$

A key point we see here is the averaging effect coming from the action

$$\varepsilon \int_{-\varepsilon^{-1}t}^{0} L(\eta(r), \dot{\eta}(r))\, dr.$$

The curve η travels in the periodic environment for a long period of time t/ε, and intuitively, it should be able to see the repeated structure. We then take average of the action and obtain the large time average heuristically. This large time average gives the homogenization result. Of course, this discussion is purely heuristic, but it is rather important to see the key point before giving rigorous treatments.

In the following, we use the optimal control formula (8.4) together with deep understanding of the metric distance to prove Theorem 8.1. The metric distance is similar to the minimal cost $h_t(x,y)$ defined in Subsection 4.3.5. The only difference is that the metric distance is defined in $\mathbb{R}^n \times \mathbb{R}^n$, not $\mathbb{T}^n \times \mathbb{T}^n$. We give some preparations in the next section.

8.1. Preliminaries and the metric problem

We assume the setting of Theorem 8.1.

8.1.1. A topological lemma. The following is a topological lemma on how to equally divide a continuous curve.

Lemma 8.2. *Let $m \in \mathbb{N}$ and $\xi : [0,1] \to \mathbb{R}^m$ be a continuous path. Then, there is a collection of disjoint intervals $\{[a_i, b_i]\}_{1\le i\le k} \subset [0,1]$ with $k \le \frac{m+1}{2}$ such that*

$$\sum_{i=1}^{k} (\xi(b_i) - \xi(a_i)) = \frac{\xi(1) - \xi(0)}{2}.$$

This is basically a generalized version of the intermediate value theorem in multi dimensions. The curve ξ travels $\xi(1) - \xi(0)$ in a unit amount of time. And Lemma 8.2 tells us that we can find a collection of disjoint time intervals $\{[a_i, b_i]\}_{1\le i\le k} \subset [0,1]$ such that the through these time intervals, the curve travels one half of the total amount, that is, $\frac{\xi(1)-\xi(0)}{2}$. It is also important that the number of these intervals is at most $\frac{m+1}{2}$. Note that we talk about the amount of travel in terms of vectors in $\mathbb{R}^m$ here.

Proof. Consider the unit sphere $\mathcal{S}^m$ in $\mathbb{R}^{m+1}$. For $x = (x_1, \ldots, x_{m+1}) \in \mathcal{S}^m$, we of course have

$$\sum_{i=1}^{m+1} x_i^2 = 1.$$

We define a map $f : \mathcal{S}^m \to \mathbb{R}^m$ as following. For each $x \in \mathcal{S}^m$, take a partition $0 = t_0 \le t_1 \le \ldots \le t_{m+1} = 1$ such that

$$t_i - t_{i-1} = x_i^2 \quad \text{for } 1 \le i \le m+1.$$

In other words, for $1 \le i \le m+1$,

$$t_i = \sum_{j=1}^{i} x_j^2.$$

Denote by

$$f(x) = \sum_{i=1}^{m+1} \operatorname{sign}(x_i)(\xi(t_i) - \xi(t_{i-1})).$$

Note that if $x_i = 0$, then $t_{i-1} = t_i$. Thus, f is well-defined. It is not hard to see that $f \in C(\mathcal{S}^m, \mathbb{R}^m)$ and f is odd, that is,

$$f(x) = -f(-x) \quad \text{for } x \in \mathcal{S}^m.$$

Thus, by the Borsuk–Ulam theorem, there exists $\bar{x} \in \mathcal{S}^m$ such that

$$f(\bar{x}) = f(-\bar{x}) \quad \Longrightarrow \quad f(\bar{x}) = 0.$$

Here, $\bar{x}$ is called an antipodal point. Without loss of generality, we assume that $\bar{x}$ has at most $\frac{m+1}{2}$ positive coordinates. Then, the collection of disjoint intervals $[t_{i-1}, t_i]$ with $\bar{x}_i > 0$ is exactly what we need. The proof is complete. $\square$

8.1.2. A priori estimates and simplifications. By the usual comparison principle, we have

$$\|u_t^\varepsilon\|_{L^\infty(\mathbb{R}^n \times [0,\infty))} + \|Du^\varepsilon\|_{L^\infty(\mathbb{R}^n \times [0,\infty))} \le C_0.$$

Here, $C_0 > 0$ is a constant depending only on H and $\|Dg\|_{L^\infty(\mathbb{R}^n)}$. In particular, we see that the values of $H(y,p)$ for $|p| > C_0$ are irrelevant. By modifying $H(y,p)$ for $|p| > 2C_0 + 1$ if needed, we assume further that H grows quadratically in p, that is,

$$\frac{1}{2}|p|^2 - K_0 \le H(y,p) \le \frac{1}{2}|p|^2 + K_0 \quad \text{for all } (y,p) \in \mathbb{T}^n \times \mathbb{R}^n, \tag{8.5}$$

for some $K_0 > 1$. Let $L(y,v)$ be the Lagrangian (Legendre transform) of the Hamiltonians $H(y,p)$. It is clear that

$$\frac{1}{2}|v|^2 - K_0 \le L(y,v) \le \frac{1}{2}|v|^2 + K_0 \quad \text{for all } (y,v) \in \mathbb{T}^n \times \mathbb{R}^n. \tag{8.6}$$

The quadratic growth of both H and L helps us control various bounds in a more intuitive way later on. Besides, as our estimates are independent of the smoothness of H and L, by approximations, we may assume further that

$$H, L \in C^k(\mathbb{T}^n \times \mathbb{R}^n)$$

for some $k \geq 2$.

8.1.3. The metric distance. For $x, y \in \mathbb{R}^n$ and $t > 0$, denote by

$$m(t,x,y) = \inf\left\{\int_0^t L(\eta(s),\dot\eta(s))\,ds \,:\, \eta \in \mathrm{AC}\,([0,t],\mathbb{R}^n), \eta(0)=x, \eta(t)=y\right\}.$$

Here, $m(t,x,y)$ is the minimum cost to travel from x to y in a given time $t > 0$. We say that $m(t,x,y)$ is the metric distance from x to y in time t. Again, the metric distance is similar to the minimal cost $h_t(x,y)$ defined in Subsection 4.3.5. The only difference is that the metric distance is defined in $\mathbb{R}^n \times \mathbb{R}^n$, not $\mathbb{T}^n \times \mathbb{T}^n$.

The homogenized (large time average) metric is

$$\overline{m}(t,x,y) = \lim_{k\to\infty} \frac{1}{k} m(kt,kx,ky). \tag{8.7}$$

In fact,

$$\overline{m}(t,x,y) = t\overline{L}\left(\frac{y-x}{t}\right),$$

where $\overline{L}$ is the Lagrangian (Legendre transform) of the effective Hamiltonian H. In particular, for $s > 0$,

$$\overline{m}(st,sx,sy) = st\overline{L}\left(\frac{y-x}{t}\right) = s\overline{m}(t,x,y).$$

Thus, $\overline{m}$ is positively homogeneous of degree one. Some basic properties of m are collected in the following lemma. To make the presentation self-contained, we prove that (8.7) holds also in this lemma.

Lemma 8.3. *Assume* (8.5)–(8.6). *Then, there exists $C > 0$, a universal constant depending only on L and n, such that we have the following properties.*

(a) *m is subadditive, that is, for $x, y, z \in \mathbb{R}^n$ and $t, s > 0$,*

$$m(t,x,y) + m(s,y,z) \geq m(t+s,x,z).$$

(b) *m is periodic, that is, $x, y \in \mathbb{R}^n$, $w \in \mathbb{Z}^n$, and $t > 0$,*

$$m(t,x+w,y+w) = m(t,x,y).$$

(c) *For $t > 0$, and $|y| \leq Ct$,*

$$m(2t,0,2y) \leq 2m(t,0,y) + C.$$

Generally speaking, for $r, l > 0$,

$$m((r+l)t,0,(r+l)y) \leq m(rt,0,ry) + m(lt,0,ly) + C.$$

(d) *The convergence* (8.7) *holds, that is, for given $x, y \in \mathbb{R}^n$ and $t > 0$, there exists $\overline{m}(t,x,y) \in \mathbb{R}$ such that,*

$$\overline{m}(t,x,y) = \lim_{k\to\infty} \frac{1}{k} m(kt,kx,ky).$$

Moreover, $\overline{m}$ is positively homogeneous of degree one.

(e) *For $t > 0$, and $|y| \le Ct$,*

$$\overline{m}(t,0,y) \le m(t,0,y) + C. \tag{8.8}$$

(f) *For $\varepsilon, t > 0$, and $|y| \le Ct$,*

$$\overline{m}(t,0,y) \le \varepsilon m\left(\frac{t}{\varepsilon}, 0, \frac{y}{\varepsilon}\right) + C\varepsilon. \tag{8.9}$$

Proof. We note that (a) is just the usual triangle inequality and its proof is clear from the definition of m. Also, (b) follows directly from the fact that L is $\mathbb{Z}^n$-periodic in y.

Let us now prove (c). Write

$$y = [y] + \tilde{y},$$

where $[y] \in \mathbb{Z}^n$ is the integer part of y, and $\tilde{y} \in [0,1)^n$. By (a) and (b),

$$\begin{aligned} m(2t,0,2y) &\le m(t,0,y) + m(t,y,2y) \\ &= m(t,0,y) + m(t,\tilde{y},y+\tilde{y}). \end{aligned}$$

By Theorem 4.34, as $\tilde{y} \in [0,1)^n$,

$$|m(t,0,y) - m(t,\tilde{y},y+\tilde{y})| \le C|\tilde{y}| \le C.$$

We thus obtain

$$m(2t,0,2y) \le 2m(t,0,y) + C.$$

Similarly, for $r, l > 0$,

$$m((r+l)t,0,(r+l)y) \le m(rt,0,ry) + m(lt,0,ly) + C.$$

Let us now prove (d). It is enough to consider the case $x = 0$ as the general case follows in a similar manner. Fix $y \in \mathbb{R}^n$ and $t > 0$. Then, we can find $C > 0$ such that $|y| \le Ct$. Let $\phi : [0,\infty) \to \mathbb{R}$ be such that, for $l \ge 0$,

$$\phi(l) = m(lt,0,ly) + C.$$

Then, in light of (c), ϕ is subadditive, that is, for $l, r \ge 0$,

$$\phi(l+r) \le \phi(l) + \phi(r).$$

Thanks to Fekete's lemma,

$$\lim_{k\to\infty} \frac{\phi(k)}{k} = \inf_{l>0} \frac{\phi(l)}{l}.$$

See Appendix D for an elementary proof of Fekete's lemma. Note that, by (8.6), $\phi(l) \geq -K_0 l + C$, which yields that the right hand side of the above equality is finite. We conclude the proof of (8.7) and also that the limit is finite. It is quite clear that $\overline{m}$ is positively homogeneous of degree one as for $s > 0$,

$$\begin{aligned}\overline{m}(st, sx, sy) &= \lim_{k\to\infty} \frac{1}{k} m(kst, ksx, ksy)\\ &= s \lim_{k\to\infty} \frac{1}{ks} m(kst, ksx, ksy) = s\overline{m}(t, x, y).\end{aligned}$$

Next, we prove (e). From (c), we see that

$$\frac{1}{2}\left(m(2t, 0, 2y) + C\right) \leq m(t, 0, y) + C.$$

By iterations, for $k \in \mathbb{N}$,

$$\frac{1}{2^k}\left(m(2^k t, 0, 2^k y) + C\right) \leq m(t, 0, y) + C.$$

Let $k \to \infty$ to deduce

$$\overline{m}(t, 0, y) \leq m(t, 0, y) + C.$$

Finally, we prove (f). By (e),

$$\overline{m}\left(\frac{t}{\varepsilon}, 0, \frac{y}{\varepsilon}\right) \leq m\left(\frac{t}{\varepsilon}, 0, \frac{y}{\varepsilon}\right) + C.$$

As $\overline{m}$ is positively homogeneous of degree one, we imply

$$\overline{m}(t, 0, y) \leq \varepsilon m\left(\frac{t}{\varepsilon}, 0, \frac{y}{\varepsilon}\right) + C\varepsilon.$$

$\square$

Next, we show that m is superadditive, which is harder to obtain in general.

Lemma 8.4. *Assume* (8.5)–(8.6). *Then, for $t > n$ and $y \in \mathbb{R}^n$ with $|y| \leq Ct$,*

$$2m(t, 0, y) \leq m(2t, 0, 2y) + C. \tag{8.10}$$

In particular,

$$m(t, 0, y) \leq \overline{m}(t, 0, y) + C, \tag{8.11}$$

and, for $\varepsilon > 0$,

$$\varepsilon m\left(\frac{t}{\varepsilon}, 0, \frac{y}{\varepsilon}\right) \leq \overline{m}(t, 0, y) + C\varepsilon. \tag{8.12}$$

Here, $C > 0$ is a universal constant depending only on L and n.

Proof. It is enough to prove (8.10). The proofs of (8.11) and (8.12) then follow in a similar manner like those in the end of the proof of Lemma 8.3. Hereafter, $C > 0$ represents a universal constant depending only on L and n. By considering $\alpha(s) = sy/t$ for $s \in [0, 2t]$, we deduce that $m(2t, 0, 2y) \le Ct$. Thanks to Theorem 2.23, there exists $\gamma : [0, 2t] \to \mathbb{R}^n$ with $\gamma(0) = 0$, $\gamma(2t) = 2y$, and $\gamma \in C^k$ such that

$$m(2t, 0, 2y) = \int_0^{2t} L(\gamma(s), \dot\gamma(s))\, ds \le Ct. \tag{8.13}$$

Let $\xi(s) = (\gamma(s), s)$ for $s \in [0, 2t]$, that is, we add the time variable also to the curve ξ. By Lemma 8.2, we are able to find a collection of disjoint intervals $\{[a_i, b_i]\}_{1 \le i \le k} \subset [0, 2t]$ with $k \le \frac{n+2}{2}$ such that

$$\sum_{i=1}^{k} (\xi(b_i) - \xi(a_i)) = \frac{\xi(2t) - \xi(0)}{2} = (y, t).$$

Rearranging and shifting γ on $\{[a_i, b_i]\}_{i=1}^k$ in a periodic way in space to get $\tilde\gamma : (0, t) \to \mathbb{R}^n$ such that, for $t_0 = 0$, $t_j = \sum_{i=1}^{j} (b_i - a_i)$ for $1 \le j \le k$,

- $\tilde\gamma(0^+) \in [0, 1]^n$;
- $\tilde\gamma|_{(t_{j-1}, t_j)}$ is a periodic shift of $\gamma|_{(a_j, b_j)}$ for $1 \le j \le k$;
- for $1 \le j \le k-1$, $\tilde\gamma(t_j^+) - \tilde\gamma(t_j^-) \in [0, 1]^n$, which gives
$$\left|\tilde\gamma(t_j^+) - \tilde\gamma(t_j^-)\right| \le \sqrt{n};$$
- $$\sum_{i=1}^{k} (\tilde\gamma(t_i^-) - \tilde\gamma(t_{i-1}^+)) = y.$$

Set $\tilde\gamma(0^-) = 0$, and $\tilde\gamma(t^+) = y$. We now use $\tilde\gamma$ to create $\eta \in \mathrm{AC}([0, t], \mathbb{R}^n)$ with $\eta(0) = 0, \eta(t) = y$, and

$$\int_0^t L(\eta(s), \dot\eta(s))\, ds \le \int_0^t L(\tilde\gamma(s), \dot{\tilde\gamma}(s))\, ds + C, \tag{8.14}$$

that is, we are off by at most a constant cost. We create η by using $\tilde\gamma$ and connectors. Here, all connectors are straight lines with constant velocity for simplicity (other options also work).

If $t < n$, then we simply let η be the connector connecting 0 to y, that is, $\eta(s) = sy/t$ for $s \in [0, t]$. It is clear that (8.14) holds.

Let us now consider the case where $t \ge n$. By (8.13), there exists $d \in \{0, 1, \ldots, [t] - 1\}$ such that

$$\int_d^{d+1} L(\tilde\gamma(s), \dot{\tilde\gamma}(s))\, ds \le C.$$

By the bound (8.6), we yield

$$\int_d^{d+1} |\dot{\tilde{\gamma}}(s)|^2\, ds \le C. \tag{8.15}$$

Let us now rescale $\tilde{\gamma}$ on $[d, d+1]$ to save an amount of time $1/2$. Denote by

$$\alpha(s) = \begin{cases} \tilde{\gamma}(s) & \text{for } 0 \le s \le d, \\ \tilde{\gamma}(d + 2(s-d)) & \text{for } d \le s \le d + \frac{1}{2}, \\ \tilde{\gamma}(s + \frac{1}{2}) & \text{for } d + \frac{1}{2} \le s \le t - \frac{1}{2}. \end{cases}$$

Thanks to (8.15),

$$\left| \int_0^t L(\tilde{\gamma}(s), \dot{\tilde{\gamma}}(s))\, ds - \int_0^{t-\frac{1}{2}} L(\alpha(s), \dot{\alpha}(s))\, ds \right| \le C. \tag{8.16}$$

The main point here is that α saves an amount of time $1/2$, and this amount of time can be used to go along the $k+1$ connectors suitably.

We next create $k+1$ connectors, each takes an amount of time $1/(2k+2)$ connecting $\tilde{\gamma}(t_j^-)$ to $\tilde{\gamma}(t_j^+)$ for $0 \le j \le k$. We then glue the pieces of α together with these $k+1$ connectors to get the desired path η. See Figure 1. In light of (8.16), (8.14) holds true. Combining (8.13) and (8.14), we arrive

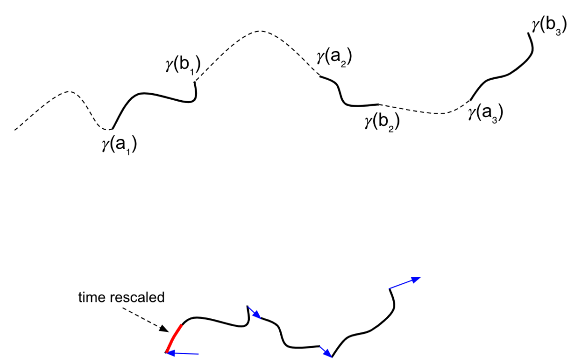

Figure 1. Formation of the curve η

at

$$2m(t, 0, y) \le m(2t, 0, 2y) + C.$$

□

8.2. Proof of Theorem 8.1

Proof of Theorem 8.1. We combine (8.8), (8.9), (8.11), and (8.12) to yield, for $\varepsilon, s > 0$ and $y \in \mathbb{R}^n$ with $|y| \leq Cs$,

$$|m(s,0,y) - \overline{m}(s,0,y)| \leq C, \tag{8.17}$$

and

$$\left|\varepsilon m\left(\frac{s}{\varepsilon}, 0, \frac{y}{\varepsilon}\right) - \overline{m}(s,0,y)\right| \leq C\varepsilon. \tag{8.18}$$

By suitable scalings and translations, it suffices to obtain the result for $(x,t) = (0,1)$. As $g \in \mathrm{Lip}\,(\mathbb{R}^n)$,

$$|g(x)| \leq |g(x) - g(0)| + |g(0)| \leq C(|x|+1) \qquad \text{for all } x \in \mathbb{R}^n. \tag{8.19}$$

Recall that the optimal control formula (8.4) gives us that

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

Thanks to (8.6) and the Jensen inequality,

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

Combining the above with (8.19) and the optimal control formula, we yield that the infimum in the optimal control formula only happens when

$$\varepsilon \left|\eta(-\varepsilon^{-1})\right| \leq C$$

for $C > 0$ depending only on L, $\|Dg\|_{L^\infty(\mathbb{R}^n)}$, and n.

Therefore, we can write

$$\begin{aligned} u^\varepsilon(0,1) &= \inf_{\substack{\eta(0)=0,\\ \varepsilon\left|\eta(-\varepsilon^{-1})\right|\leq C}} \left\{ g\left(\varepsilon\eta\left(-\varepsilon^{-1}\right)\right) + \varepsilon \int_{-\varepsilon^{-1}}^{0} L(\eta(t),\dot\eta(t))\,dt \right\} \\ &= \inf_{|y|\leq C} (g(y) + \varepsilon m(\varepsilon^{-1}, \varepsilon^{-1}y, 0)) \\ &= \inf_{|y|\leq C} (g(y) + \overline{m}(1,0,-y)) + O(\varepsilon) \\ &= u(0,1) + O(\varepsilon). \end{aligned}$$

We used (8.18) in the second last equality. The proof is complete. □

8.3. 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. We note that there are many such examples, and we choose one that is relatively simple to present here.

Proposition 8.5. *Assume that $n = 1$ and*

$$H(y,p) = \frac{p^2}{2} + V(y) \qquad \text{for } (y,p) \in \mathbb{T} \times \mathbb{R}$$

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^ε be the solution to (8.2)*. Let u be the solution to* (8.3)*. Then, u^ε converges locally uniformly to $u \equiv 0$ on $\mathbb{R} \times [0,\infty)$ as $\varepsilon \to 0$. Furthermore, for $\varepsilon \in (0,1)$,*

$$u^\varepsilon(0,1) \geq \frac{\varepsilon}{6}. \tag{8.20}$$

Proof. Note that $\overline{H}(0) = 0$, and thus, $u \equiv 0$. We already know that u^ε converges locally uniformly to $u \equiv 0$ on $\mathbb{R} \times [0,\infty)$ as $\varepsilon \to 0$. We just need to prove (8.20).

As usual, the optimal control formula gives

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

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

Case 1. If $\eta([0,3^{-1}]) \subset [-3^{-1}, 3^{-1}]$, then

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

Case 2. $\eta([0,3^{-1}]) \not\subseteq [-3^{-1}, 3^{-1}]$. Then, without loss of generality, we assume that there exists $t \in (0, 3^{-1})$ such that $\eta(t) = 3^{-1}$. We use Jensen's inequality to deduce that

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

The proof is complete. □

8.4. Open problems

The main remaining open problem in the quantitative theory is to find optimal convergence rates for periodic homogenization of nonconvex Hamilton-Jacobi equations, where all methods and ideas from stable norms and first passage percolations cease to work. The first nonconvex situation with $O(\varepsilon)$ convergence rate was obtained in [**TY21**] for

$$H(y,p) = \max\{|p| - 1, 1 - |p|\} + V(y) \qquad \text{for all } (y,p) \in \mathbb{T}^n \times \mathbb{R}^n$$

with $\operatorname{osc}(V) = \max V - \min V \geq 1$. One possible strategy to handle the nonconvex case is to first identify the shape of the effective Hamiltonian, then design customized strategies based on the game theory interpretation.

Question 5. *Identify and prove the optimal rate of convergence of u^ε to u in the general nonconvex setting.*

Question 6. *Theorem 8.1 gives us that $\|u^\varepsilon - u\|_{L^\infty(\mathbb{R}^n\times[0,\infty))} \leq C\varepsilon$. Identify any possible pattern of $\frac{u^\varepsilon - u}{\varepsilon}$ as $\varepsilon \to 0+$.*

Question 7. *Study the optimal rate of convergence in the convex setting when H has multiscale nature, for example,*

$$H = H\left(x, \frac{x}{\varepsilon}, \frac{t}{\varepsilon}, \frac{u^\varepsilon}{\varepsilon}\right).$$

8.5. References

(1) Much of the content of this chapter is based on Tran, Yu [**TY21**].

(2) Proposition 8.5 is based on an example given in Mitake, Tran, Yu [**MTY19**].

(3) We give a minimalistic review of the PDE literature playing major roles in finding the convergence rate in the periodic setting.

For the general nonconvex setting, the best known convergence rate is $O(\varepsilon^{1/3})$ obtained by Capuzzo-Dolcetta and Ishii [**CDI01**]. Although the result in [**CDI01**] concerns static Hamilton-Jacobi equations, the extension to the Cauchy problem is quite standard (see, e.g., [**Tra21**]).

For convex Hamilton-Jacobi equations, by using weak KAM methods, the lower bound $u^\varepsilon - u \geq -C\varepsilon$ was proved by Mitake, Tran, and Yu [**MTY19**] for all dimensions. When $n = 2$, for positive homogeneous Hamiltonian, the upper bound $u^\varepsilon - u \leq C\varepsilon$ was also derived via the classical Aubry-Mather theory, which however heavily relies on the two dimensional topology. After [**MTY19**], the major open problem was whether the upper bound $u^\varepsilon - u \leq C\varepsilon$ always holds in higher dimensions ($n \geq 3$), which was completely unclear although some conditional higher dimensional results in [**MTY19**] sort of imply that the upper bound should hold in "generic" situations. See also Jing, Tran, Yu [**JTY20**], and Tu [**Tu**].

Then, Cooperman discovered that closely related convergence rate results have been established in the context of first passage percolation in the 1990s by Alexander [**Ale90, Ale97**]. By adjusting the methods there, Cooperman was able to obtain in [**Coo22**] a near optimal convergence rate $|u^\varepsilon(x,t) - u(x,t)| \leq C\varepsilon\log(C + \varepsilon^{-1}t)$ when $n \geq 3$, which was a very surprising result for people in the PDE community.

Later on, while studying finer properties of effective fronts, the authors discovered an old result concerning the convergence rate of stable norms in metric geometry by Burago [**Bur92**] also in the 1990s that is basically equivalent to the optimal convergence rate $O(\varepsilon)$ in the homogenization of static Hamilton-Jacobi equations. Lemma 8.2 proved by Burago and Perelman played the key role there. By using this crucial lemma, we obtained Theorem 8.1, which concludes the study of this whole program in the convex setting.

Chapter 9

Large time behavior for Hamilton-Jacobi equations in the torus

In this chapter, we always consider a given Hamiltonian $H : \mathbb{T}^n \times \mathbb{R}^n \to \mathbb{R}$ that satisfies

$$
\begin{cases} H \in C^k(\mathbb{T}^n \times \mathbb{R}^n) \text{ for some } k \geq 2, \\ D^2_{pp}H(y,p) > 0 \text{ for all } (y,p) \in \mathbb{T}^n \times \mathbb{R}^n, \\ \lim_{|p|\to\infty} \min_{y\in\mathbb{T}^n} \frac{H(y,p)}{|p|} = +\infty. \end{cases} \tag{9.1}
$$

Let L be the corresponding Lagrangian (the Legendre transform of H). Then, L satisfies

$$
\begin{cases} L \in C^k(\mathbb{T}^n \times \mathbb{R}^n), \\ D^2_{vv}L(y,v) > 0 \text{ for all } (y,v) \in \mathbb{T}^n \times \mathbb{R}^n, \\ \lim_{|v|\to\infty} \min_{y\in\mathbb{T}^n} \frac{L(y,v)}{|v|} = +\infty. \end{cases} \tag{9.2}
$$

The main object in this chapter is the following Cauchy problem

$$
\begin{cases} u_t + H(x, Du) = 0 & \text{in } \mathbb{T}^n \times (0,\infty), \\ u(x,0) = g(x) & \text{on } \mathbb{T}^n. \end{cases} \tag{9.3}
$$

Here, $g \in C(\mathbb{T}^n)$ is the given initial data, and $u : \mathbb{T}^n \times [0,\infty) \to \mathbb{R}$ is the unknown. Our main goal in this chapter is to study the large time behavior of u, that is, $\lim_{t\to\infty} u(x,t)$ after appropriate normalizations.

Heuristically, it is natural to expect that, as $t \to \infty$,

$$
u(x,t) \approx v(x) - ct,
$$

which can be considered as an ansatz. Plug this expansion into (9.3), we see that

$$H(x, Dv(x)) = c \qquad \text{in } \mathbb{T}^n.$$

Thus, we see that $c = c[0]$, and we again need the cell problem at $p = 0$, that is,

$$H(x, Dv(x)) = \overline{H}(0) = c[0] \qquad \text{in } \mathbb{T}^n. \tag{9.4}$$

Here, $c[0] = \overline{H}(0) \in \mathbb{R}$ is the unique constant so that (9.4) has a viscosity solution as discussed in the previous chapters. Sometimes, $c[0] = \overline{H}(0)$ is also called the ergodic constant in the literature.

9.1. Large time behavior

Let us state the main result of this chapter.

Theorem 9.1. *Assume* (9.1). *Let $g \in C(\mathbb{T}^n)$ be a given initial data, and u be the viscosity solution to* (9.3). *Then, $u(x,t) + c[0]t \to v$ uniformly on $\mathbb{T}^n$ as $t \to \infty$, where v is a viscosity solution to* (9.4).

To date, there have been many different proofs of this important large time behavior result. We present here a proof following Fathi [**Fat**].

We always assume the settings of Theorem 9.1 in this section. We first need the following important preparation lemma.

Lemma 9.2. *Fix $\varepsilon > 0$. Then, there exists $t(\varepsilon) > 0$ such that, for each $t > t(\varepsilon)$, if $Du(x,t)$ exists, then*

$$c[0] - \varepsilon \le H(x, Du(x,t)) \le c[0] + \varepsilon.$$

Proof. Fix $\varepsilon > 0$. Denote by

$$W_\varepsilon = \{(x,v) \in \mathbb{T}^n \times \mathbb{R}^n \, : \, c[0] - \varepsilon < H \circ \mathcal{L}(x,v) < c[0] + \varepsilon\}.$$

Then, W_ε is a neighborhood of $\widetilde{\mathcal{M}}_0$ thanks to Lemma 5.6. By Lemma 5.21, there exists $t(\varepsilon) > 0$ such that for any minimizing curve $\gamma : [0,t] \to \mathbb{T}^n$ with $t > t(\varepsilon)$ then we can find $t' \in [0,t]$ with

$$\big(\gamma(t'), \dot\gamma(t')\big) \in W_\varepsilon.$$

By the conservation of energy, we imply further that

$$H \circ \mathcal{L}\left(\gamma(s), \dot\gamma(s)\right) \in [c[0] - \varepsilon, c[0] + \varepsilon] \qquad \text{for all } s \in [0,t]. \tag{9.5}$$

Now, for $t > t(\varepsilon)$, if $Du(x,t)$ exists, then we are able to find a minimizing curve $\gamma : [0,t] \to \mathbb{T}^n$ with $\gamma \in C^k([0,t])$, $\gamma(t) = x$, and

$$\begin{cases} u(x,t) = \int_0^t L(\gamma(s), \dot\gamma(s))\, ds + g(\gamma(0)), \\ Du(x,t) = D_v L(\gamma(t), \dot\gamma(t)). \end{cases}$$

By (9.5), we conclude that

$$c[0] - \varepsilon \leq H(x, Du(x,t)) \leq c[0] + \varepsilon.$$

□

Proof of Theorem 9.1. Without loss of generality, we assume $c[0] = 0$. By approximations, we may also assume $g \in \mathrm{Lip}\,(\mathbb{T}^n)$. We divide the proof into several steps.

Step 1. Let $v_0 \in C(\mathbb{T}^n)$ be a solution to (9.4). Pick $C = \|v_0\|_{L^\infty(\mathbb{T}^n)} + \|g\|_{L^\infty(\mathbb{T}^n)}$. Then,

$$v_0 - C \leq g \leq v_0 + C.$$

Note that $v_0 \pm C$ are separable solutions to (9.3) with initial data $v_0 \pm C$, respectively. By the usual comparison principle to (9.3),

$$v_0(x) - C \leq u(x,t) \leq v_0(x) + C \qquad \text{for all } (x,t) \in \mathbb{T}^n \times [0,\infty). \tag{9.6}$$

Since $g \in \mathrm{Lip}\,(\mathbb{T}^n)$, there exists $C > 0$ such that

$$\|u_t\|_{L^\infty(\mathbb{T}^n\times[0,\infty))} + \|Du\|_{L^\infty(\mathbb{T}^n\times[0,\infty))} \leq C. \tag{9.7}$$

Step 2. By (9.6) and (9.7), we use the Arzelà-Ascoli theorem to find a sequence $\{t_k\} \to \infty$ such that

$$T_{t_k} g(x) = u(x,t_k) \to u_\infty(x) \qquad \text{uniformly on } \mathbb{T}^n \text{ as } t_k \to \infty.$$

By Lemma 9.2, for each $\varepsilon > 0$, there exists $t(\varepsilon) > 0$ such that, for $t_k > t(\varepsilon)$,

$$H(x, Du(x,t_k)) \leq c[0] + \varepsilon = \varepsilon \qquad \text{for a.e. } x \in \mathbb{T}^n.$$

In particular, $u(x,t_k)$ is a viscosity subsolution to

$$H(x, Du(x,t_k)) \leq \varepsilon \qquad \text{in } \mathbb{T}^n.$$

Let $t_k \to \infty$ and $\varepsilon \to 0$ in this order to yield that u_∞ is a viscosity subsolution to

$$H(x, Du_\infty) \leq 0 \qquad \text{in } \mathbb{T}^n. \tag{9.8}$$

Step 3. As u_∞ is a subsolution to (9.8), we see that

$$T_t u_\infty(x) \leq u_\infty(x) \qquad \text{for all } t \geq 0. \tag{9.9}$$

Without loss of generality, we assume $t_{k+1} - t_k \to \infty$ as $k \to \infty$. Denote by $s_k = t_{k+1} - t_k$. We claim that

$$T_{s_k} u_\infty \to u_\infty \qquad \text{uniformly on } \mathbb{T}^n \text{ as } k \to \infty. \tag{9.10}$$

Indeed, by the comparison principle and the triangle inequality

$$\begin{aligned}\|T_{s_k}u_\infty - u_\infty\|_{L^\infty} &\le \|T_{t_{k+1}}g - T_{s_k}u_\infty\|_{L^\infty} + \|T_{t_{k+1}}g - u_\infty\|_{L^\infty}\\ &= \|T_{s_k}\circ T_{t_k}g - T_{s_k}u_\infty\|_{L^\infty} + \|T_{t_{k+1}}g - u_\infty\|_{L^\infty}\\ &\le \|T_{t_k}g - u_\infty\|_{L^\infty} + \|T_{t_{k+1}}g - u_\infty\|_{L^\infty} \to 0\end{aligned}$$

as $k \to \infty$. Hence, (9.10) is valid.

Step 4. Thanks to (9.9) and (9.10),

$$T_t u_\infty(x) = u_\infty(x) \qquad \text{for all } t \ge 0. \tag{9.11}$$

We now only need to use the comparison principle to show that

$$T_t u_\infty \to u_\infty \qquad \text{uniformly on } \mathbb{T}^n \text{ as } t \to \infty. \tag{9.12}$$

Indeed, for $t \ge t_k$, we use (9.11) and the usual comparison principle to get

$$\begin{aligned}\|T_t g - u_\infty\|_{L^\infty} &= \|T_{t-t_k}\circ T_{t_k}g - T_{t-t_k}u_\infty\|_{L^\infty}\\ &\le \|T_{t_k}g - u_\infty\|_{L^\infty},\end{aligned}$$

which gives (9.12). The proof is complete. □

9.2. A nonconvergence example

It turns out that the strict (or uniform) convexity of the Hamiltonian is really needed in order to get the large time behavior result. We now give an example to show that $u(x,t)$ does not converge as $t \to \infty$ if H is not strictly convex in p.

Example 9.3. Consider the following Hamilton-Jacobi equation in one dimension

$$u_t + |u_x + a| - a = 0 \qquad \text{in } \mathbb{T} \times (0,\infty). \tag{9.13}$$

Here, $a > 0$ is a given constant. The Hamiltonian $H(x,p) = |p+a| - a$ is convex, but not strictly convex in this case.

In this setting, (9.4) becomes

$$|v_x + a| - a = c[0] \qquad \text{in } \mathbb{T}.$$

We see that $c[0] = 0$, and $v = C$ for any given constant $C \in \mathbb{R}$ is a solution to the above.

Set

$$u(x,t) = \frac{a}{4\pi}\sin(2\pi(x-t)) \qquad \text{for } (x,t) \in \mathbb{T} \times [0,\infty).$$

It is clear that, for $(x,t) \in \mathbb{T} \times [0,\infty)$,

$$u_t(x,t) = -\frac{a}{2}\cos(2\pi(x-t)) \qquad \text{and} \qquad u_x(x,t) = \frac{a}{2}\cos(2\pi(x-t)).$$

Hence, u is a solution to (9.13) with initial data

$$g(x) = u(x,0) = \frac{a}{4\pi}\sin(2\pi x).$$

However, as $t \to \infty$, $u(x,t)$ does not converge to any limit.

9.3. Maximal subsolutions and a different definition of the Aubry set

We only focus on the cell problem (9.4) here. We recall that $h_t(x,y)$ is the minimal cost it takes to travel from x to y in a given fixed amount of time t corresponding to the given Lagrangian L. More specifically, as defined in (4.13),

$$h_t(x,y) = \inf_{\substack{\gamma\in \mathrm{AC}\,([0,t],\mathbb{T}^n)\\ \gamma(0)=x,\gamma(t)=y}} \int_0^t L(\gamma(s),\dot\gamma(s))\,ds.$$

Definition 9.4. For $x, y \in \mathbb{T}^n$, denote by

$$d(x,y) = \inf\Big\{\int_0^t L(\gamma(s),\dot\gamma(s))\,ds + c[0]t\ :\ t>0, \gamma\in \mathrm{AC}\,([0,t],\mathbb{T}^n), \gamma(0)=x, \gamma(t)=y\Big\}.$$

From the definition of d, we see that

$$d(x,y) = \inf\left\{h_t(x,y) + c[0]t\ :\ t>0\right\}.$$

Besides, it is worth to note that d is different from the Peierls barrier h, where

$$h(x,y) = \liminf_{t\to\infty}\left[h_t(x,y) + c[0]t\right].$$

In particular, we observe that, for $x, y \in \mathbb{T}^n$,

$$d(x,y) \le h(x,y).$$

Theorem 9.5 (Properties of d)**.** *Assume* (9.1)*. Then, we have the following properties of d.*

(a) *For $x, y \in \mathbb{T}^n$,*

$$d(x,y) = \sup\{v(y) - v(x)\ :\ v \text{ is a subsolution to (9.4)}\}.$$

(b) *For $x, y, z \in \mathbb{T}^n$, $d(x,x) = 0$, and*

$$d(x,z) \le d(x,y) + d(y,z).$$

(c) *For $x \in \mathbb{T}^n$ fixed, $y \mapsto d(x,y)$ is a subsolution to* (9.4)*, and is a solution to* (9.4) *in $\mathbb{T}^n \setminus \{x\}$.*

Proof. We first prove (a). By Lemma 3.14, for each v being a subsolution to (9.4) and $x, y \in \mathbb{T}^n$,

$$h_t(x,y) + c[0]t \geq v(y) - v(x).$$

Taking infimum over $t > 0$ and supremum over v in this order to yield

$$d(x,y) \geq \sup\{v(y) - v(x) \,:\, v \text{ is a subsolution to (9.4)}\}.$$

To conclude, we show that $y \mapsto w(y) = d(x,y)$ is a subsolution to (9.4). It is not hard to see that w is Lipschitz. Pick $y \neq x$ to be a differentiable point of w. We need to show that

$$H(y, Dw(y)) \leq c[0]. \tag{9.14}$$

By the definition of d, we see that, for $0 < r < |y-x|/2$,

$$w(y) = \inf_{z \in \partial B(y,r)} \left(w(z) + d(z,y) \right).$$

For each nonzero vector $e \in \mathbb{R}^n$, denote

$$\gamma_e(s) = y - te + se \qquad \text{for } 0 \leq s \leq t$$

for $t > 0$ sufficiently small. Then, by using this path γ_e and the above relation, we see that

$$w(y) \leq w(y - te) + \int_0^t L(y - te + se, e)\, ds + c[0]t.$$

Hence,

$$\frac{w(y) - w(y-te)}{t} \leq \frac{1}{t} \int_0^t L(y - te + se, e)\, ds + c[0].$$

Let $t \to 0^+$ in the above to deduce that

$$Dw(y) \cdot e - L(y,e) \leq c[0].$$

Maximize this inequality over $e \in \mathbb{R}^n$ to imply (9.14).

We next prove (b). The triangle inequality $d(x,z) \leq d(x,y) + d(y,z)$ is immediate from the definition of d. By part (a), we see that $d(x,x) \geq 0$. Besides, as $h_0(x,x) = 0$, we conclude that $d(x,x) = 0$.

Part (c) is also an immediate consequence of part (a). We already showed that $y \mapsto d(x,y)$ is a subsolution to (9.4). Moreover, by Perron's method and the supremum formula of d, $y \mapsto d(x,y)$ is a solution to (9.4) in $\mathbb{T}^n \setminus \{x\}$. In a more explicit way, $y \mapsto d(x,y)$ solves

$$\begin{cases} H(y, D_y d(x,y)) = c[0] & \text{in } \mathbb{T}^n \setminus \{x\}, \\ d(x,x) = 0. \end{cases}$$

$\square$

Because of the characterization of d in part (a) of the theorem above, we say that $y \mapsto d(x,y)$ is the maximal subsolution of (9.4) with given vertex x.

Theorem 9.6 (Another characterization of the Aubry set)**.** *Assume* (9.1). *Then, $x \in \mathcal{A}_0$ if and only if $y \mapsto d(x,y)$ is a solution to* (9.4) *in the whole* $\mathbb{T}^n$.

Proof. We first prove the "$\Rightarrow$" direction. Assume that $x \in \mathcal{A}_0$. Then, $h(x,x) = 0$. By Theorem 5.27, for $h^x(y) = h(x,y)$, then $h^x \in \mathcal{S}_-$ and $h^x(x) = 0$. This means that h^x is a solution to (9.4), and h^x touches $d(x,\cdot)$ from above at x. For $p \in D^-d(x,x)$, we then have $p \in D^-h^x(x)$, and thus

$$H(x,p) = c[0].$$

We hence get that $y \mapsto d(x,y)$ is a solution to (9.4).

We now prove the "$\Leftarrow$" direction. Assume $y \mapsto w(y) = d(x,y)$ is a solution to (9.4). Then, there exists a calibrated curve $\gamma : (-\infty, 0] \to \mathbb{T}^n$ with $\gamma(0) = x$. In particular, for $t > 0$,

$$w(\gamma(0)) - w(\gamma(-t)) = -d(x,\gamma(-t)) = \int_{-t}^{0} L(\gamma,\dot\gamma)\,ds + c[0]t \geq d(\gamma(-t),x).$$

Thus, we obtain

$$0 \geq d(x,\gamma(-t)) + d(\gamma(-t),x) \geq d(x,x) = 0,$$

which yields further that

$$d(x,\gamma(-t)) + d(\gamma(-t),x) = 0.$$

Take $t = 1$ and use the above equality to create a loop containing x of time at least 1 that has zero cost. Thanks to Theorem 5.24, we conclude that $x \in \mathcal{A}_0$. $\square$

We use this characterization to give a new and equivalent definition of the Aubry set.

Definition 9.7 (Another definition of the Aubry set)**.** Denote by

$$\mathcal{A}_0 = \{x \in \mathbb{T}^n \,:\, y \mapsto d(x,y) \text{ is a solution to (9.4)}\}.$$

9.4. Large time profile

In Theorem 9.1, we proved that as $t \to \infty$, $u(x,t) + c[0]t \to v$ uniformly on $\mathbb{T}^n$, where v is a viscosity solution to (9.4).

Definition 9.8 (Large time profile)**.** Assume (9.1). Let $g \in C(\mathbb{T}^n)$ be a given initial data, and u be the viscosity solution to (9.3). Denote by

$$u^\infty(x) = u^\infty[g](x) = \lim_{t\to\infty} \left(u(x,t) + c[0]t\right).$$

We say that $u^\infty = u^\infty[g]$ is the large time profile of the given initial data g. When there is no confusion, we write u^∞ in place of $u^\infty[g]$ for short.

Our goal in this section is to give a representation formula of $u^\infty = u^\infty[g]$ in terms of g and the underlying dynamics.

Proposition 9.9. *Assume* (9.1)*. Let $g \in C(\mathbb{T}^n)$ be a given initial data, and u be the viscosity solution to* (9.3)*. Let $u^\infty = u^\infty[g]$ be the corresponding large time profile. Then, for $y \in \mathcal{A}_0$,*

$$\begin{aligned} u^\infty(y) &= \min\left\{d(z,y) + g(z) \,:\, z \in \mathbb{T}^n\right\} \\ &= \sup\left\{v(y) \,:\, v \text{ is a subsolution to } (9.4) \text{ with } v \le g \text{ in } \mathbb{T}^n\right\}. \end{aligned}$$

Proof. For $y \in \mathbb{T}^n$, denote by

$$w(y) = \min\left\{d(z,y) + g(z) \,:\, z \in \mathbb{T}^n\right\}.$$

It is clear that $w \le g$ in $\mathbb{T}^n$. Besides, as $y \mapsto d(z,y)$ is a subsolution to (9.4) and H is convex in p, we yield that w is also a subsolution to (9.4). By the usual comparison principle, we imply

$$w(y) - c[0]t \le u(y,t) \qquad \text{for all } (y,t) \in \mathbb{T}^n \times [0,\infty).$$

Hence,

$$w \le u^\infty. \tag{9.15}$$

We next prove the converse inequality for $y \in \mathcal{A}_0$ to achieve that $w = u^\infty$ on $\mathcal{A}_0$. Fix $y \in \mathcal{A}_0$. Pick $z = z_y$ so that

$$w(y) = d(z,y) + g(z).$$

By the definition of d, for each $\varepsilon > 0$, there exists $t_\varepsilon > 0$ and a curve $\xi_\varepsilon \in \mathrm{AC}\left([0,t_\varepsilon],\mathbb{T}^n\right)$ with $\xi_\varepsilon(0) = z$, $\xi_\varepsilon(t_\varepsilon) = y$ such that

$$d(z,y) > \int_0^{t_\varepsilon} \Big(L(\xi_\varepsilon,\dot{\xi}_\varepsilon) + c[0]\Big)\, ds - \varepsilon.$$

Besides, as $y \in \mathcal{A}_0$, for each $k \in \mathbb{N}$, there exist $s_k \ge k$ and a loop $\delta_\varepsilon : [0,s_k] \to \mathbb{T}^n$ such that $\delta_\varepsilon(0) = \delta_\varepsilon(s_k) = y$, and

$$\int_0^{s_k} \Big(L(\delta_\varepsilon,\dot{\delta}_\varepsilon) + c[0]\Big)\, ds < \varepsilon.$$

We next use ξ_ε and δ_ε to create γ_ε as following

$$\gamma_\varepsilon(s) = \begin{cases} \xi_\varepsilon(s) & \text{for } s \in [0,t_\varepsilon], \\ \delta_\varepsilon(s - t_\varepsilon) & \text{for } s \in [t_\varepsilon, t_\varepsilon + s_k]. \end{cases}$$

Then, $\xi_\varepsilon \in \mathrm{AC}\,([0, t_\varepsilon + s_k], \mathbb{T}^n)$ with $\xi_\varepsilon(0) = z$ and $\xi_\varepsilon(t_\varepsilon + s_k) = y$. By the optimal control formula, we see that

$$\begin{aligned} u(y, t_\varepsilon + s_k) + c[0](t_\varepsilon + s_k) &\le \int_0^{t_\varepsilon + s_k} \Big(L(\xi_\varepsilon, \dot{\xi}_\varepsilon) + c[0] \Big)\, ds + g(z) \\ &\le d(z, y) + g(z) + 2\varepsilon. \end{aligned}$$

Let $k \to \infty$ and $\varepsilon \to 0$ in this order to yield

$$u^\infty(y) \le w(y). \tag{9.16}$$

Combine (9.15) and (9.16) to get $w = u^\infty$ on $\mathcal{A}_0$.

We next prove the second equality. Denote by

$$\phi(y) = \sup\{v(y) \,:\, v \text{ is a subsolution to (9.4) with } v \le g \text{ in } \mathbb{T}^n\}.$$

On the first hand, it is clear that w is a subsolution to (9.4) and $w \le g$. Thus, $w \le \phi$. On the other hand, let v be a subsolution to (9.4) with $v \le g$ in $\mathbb{T}^n$. Then,

$$v(y) - v(z) \le d(z, y) \quad \Longrightarrow \quad v(y) \le v(z) + d(z, y) \le g(z) + d(z, y).$$

Take infimum over z and supremum over v in this order in the above to yield $\phi \le w$. The proof is complete. □

Theorem 9.10. *Assume* (9.1). *Let $g \in C(\mathbb{T}^n)$ be a given initial data, and u be the viscosity solution to* (9.3). *Let $u^\infty = u^\infty[g]$ be the corresponding large time profile. For $x \in \mathbb{T}^n$, denote by*

$$w(x) = w_g(x) = \min\{d(z, x) + g(z) \,:\, z \in \mathbb{T}^n\}.$$

Then, for $x \in \mathbb{T}^n$,

$$\begin{aligned} u^\infty(x) &= \min\{d(z, y) + w_g(y) \,:\, y \in \mathcal{A}_0\} \\ &= \inf\{v(x) \,:\, v \textit{ is a solution to } (9.4) \textit{ with } v \ge w_g \textit{ in } \mathbb{T}^n\}. \end{aligned}$$

Proof. For $x \in \mathbb{T}^n$, set

$$\varphi(x) = \min\{d(y, x) + w_g(y) \,:\, y \in \mathcal{A}_0\}.$$

By the above proof, w_g is a subsolution to (9.4) and $w_g \le g$. It is clear that φ is a solution to (9.4) as $x \mapsto d(y, x)$ is a solution to (9.4) for $y \in \mathcal{A}_0$. We claim that

$$\varphi(y) = w_g(y) \qquad \text{for } y \in \mathcal{A}_0.$$

Indeed, for $y \in \mathcal{A}_0$ fixed, by the definition of φ, we already have $\varphi(y) \le w_g(y)$. On the other hand, as w_g is a subsolution to (9.4),

$$w_g(y) - w_g(z) \le d(z, y) \quad \Longrightarrow \quad w_g(y) \le d(z, y) + w_g(z),$$

which means that $\varphi(y) = w_g(y)$. Hence, we use Proposition 9.9 to get that

$$u^\infty(y) = w_g(y) = \varphi(y) \qquad \text{for } y \in \mathcal{A}_0.$$

As $\mathcal{A}_0$ is an uniqueness set of the cell problem (9.4), we deduce by the representation formula that, for $x \in \mathbb{T}^n$,

$$u^\infty(x) = \varphi(x) = \min\{d(y,x) + w_g(y) \,:\, y \in \mathcal{A}_0\}.$$

Let us finally prove the second equality. By the above, it is clear that φ is a solution to (9.4) and $\varphi \geq w_g$. Take any solution v to (9.4) with $v \geq w_g$. Then, for $x \in \mathbb{T}^n$,

$$\begin{aligned} v(x) &= \min\{d(y,x) + v(y) \,:\, y \in \mathcal{A}_0\} \\ &\geq \min\{d(y,x) + w_g(y) \,:\, y \in \mathcal{A}_0\} = \varphi(x). \end{aligned}$$

Thus, we get

$$\varphi(x) = \inf\{v(x) \,:\, v \text{ is a solution to (9.4) with } v \geq w_g \text{ in } \mathbb{T}^n\}.$$

The proof is complete. □

9.5. References

(1) Let us give a brief discussion on the literature of large time behavior for Hamilton-Jacobi equations. We first discuss the first-order equations. The first result in this direction was proved by Namah and Roquejoffre [**NR99**]. Fathi then gave a general convergence result [**Fat**] as presented in Theorem 9.1. Afterwards, Davini and Siconolfi [**DS06**] and Ishii [**Ish08**] refined and generalized the approach of Fathi, and studied the asymptotic problem for Hamilton-Jacobi equations on the the torus and on the whole space, respectively. Besides, Barles and Souganidis [**BS00a**] obtained large time behavior result by using a PDE method, which is completely different from the dynamical approach.

For the uniformly parabolic equations, Barles and Souganidis [**BS01**] proved the same large time behavior result. In this case, the strong maximum principle holds, and thus, the ergodic problem has a unique solution up to additive constants. The proof for the large-time convergence in [**BS01**] strongly depends on this fact.

For the general second-order case, in which the diffusion matrix might be degenerate, the large time behavior result was obtained by Cagnetti, Gomes, Mitake, and Tran [**CGMT15**].

(2) There are many excellent survey papers and lecture notes on large time behavior of Hamilton-Jacobi equations. See Achdou, Barles, Ishii, and Litvinov [**ABIL13**], Le, Mitake, Tran [**LMT17**] and the references therein for more detailed pictures.

(3) The nonconvergence example was given by Barles and Souganidis [**BS00b**].

(4) The characterization of the Aubry set in this chapter was done by Fathi and Siconolfi [**FS04, FS05**].

(5) The large time profile for the first-order case presented in this chapter was obtained by Davini and Siconolfi [**DS06**].

(6) For the large time profile for the second-order case, see Gomes, Mitake, and Tran [**GMT21**].

Appendix A

Notations

We list here various notations that are used in the book.

A.1. Notation for sets and spaces

- $n \in \mathbb{N}$ is often used to denote the dimensions.
- $\mathbb{R}^n = n$-dimensional real Euclidean space; $\mathbb{R} = \mathbb{R}^1$.
- e_i is the i-th vector in the canonical basis of $\mathbb{R}^n$ for $1 \le i \le n$, that is,
$$e_i = (0, \ldots, 0, 1, 0, \ldots, 0),$$
where 1 occurs in the i-th position.
- A typical point in $\mathbb{R}^n$ is often denoted by $x = (x_1, \ldots, x_n)$. Depending on different situations, we might regard x as a row vector or a column vector.
- For $x, y \in \mathbb{R}^n$ with $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n)$, write
$$x \cdot y = \sum_{i=1}^{n} x_i y_i \quad \text{and} \quad |x| = \sqrt{x \cdot x}.$$
- A typical point in $\mathbb{R}^n \times [0, \infty)$ is often denoted by
$$(x, t) = (x_1, \ldots, x_n, t),$$
where t often stands for the time variable.
- For a given real number $s \in \mathbb{R}$, denote by $[s]$ its integer part.
- $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ is the usual n-dimensional flat torus. When there is no confusion, we identify $\mathbb{T}^n$ with the unit cell $Y = [0, 1]^n$ with periodic boundary condition on Y.

- For an open set $U \subset \mathbb{R}^n$, we write ∂U to denote its boundary, and $\overline{U} = U \cup \partial U$ to denote its closure.
- For U, V open sets in $\mathbb{R}^n$, we write
$$U \subset\subset V$$
if $U \subset \overline{U} \subset V$, and $\overline{U}$ is compact, and say that U is compactly supported in V.
- For $x \in \mathbb{R}^n$ and $r > 0$, we denote by $B(x,r)$ the open ball in $\mathbb{R}^n$ with center x, radius r, that is,
$$B(x,r) = \{y \in \mathbb{R}^n : |y - x| < r\}.$$
Denote by $\overline{B}(x,r)$ the closed ball with center x, radius r, that is,
$$\overline{B}(x,r) = \{y \in \mathbb{R}^n : |y - x| \leq r\}.$$
We also write $B(x,r)$, $\overline{B}(x,r)$ as $B_r(x)$, $\overline{B}_r(x)$, respectively. When $x = 0$, we simply write $B_r = B_r(0)$, $\overline{B}_r = \overline{B}_r(0)$.

A.2. Notation for functions

Let $u : \mathbb{R}^n \to \mathbb{R}$ be a smooth function. We have some basic notions as following.

- $Du(x) = \nabla u(x) = \left(\frac{\partial u}{\partial x_1}(x), \ldots, \frac{\partial u}{\partial x_n}(x)\right)$.
- The Hessian of u at x is
$$D^2u(x) = \begin{pmatrix} \frac{\partial^2 u}{\partial x_1^2}(x) & \frac{\partial^2 u}{\partial x_1 \partial x_2}(x) & \ldots & \frac{\partial^2 u}{\partial x_1 \partial x_n}(x) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 u}{\partial x_n \partial x_1}(x) & \frac{\partial^2 u}{\partial x_n \partial x_2}(x) & \ldots & \frac{\partial^2 u}{\partial x_n^2}(x) \end{pmatrix}.$$
- The Laplacian of u at x is
$$\Delta u(x) = \mathrm{tr}(D^2u(x)) = \sum_{i=1}^{n} \frac{\partial^2 u}{\partial x_i^2}(x).$$

In this book, we use the notion $Du(x)$ instead of $\nabla u(x)$. We usually write u_{x_i} for $\frac{\partial u}{\partial x_i}$.

When u is not smooth, we have the following definition for subdifferential and superdifferential of u at x.

- The subdifferential of u at x is denoted by $D^-u(x)$, where
$$D^-u(x) = \left\{p \in \mathbb{R}^n : \liminf_{y \to x} \frac{u(y) - u(x) - p \cdot (y - x)}{|y - x|} \geq 0\right\}.$$

- The superdifferential of u at x is denoted by $D^+u(x)$, where
$$D^+u(x) = \left\{p \in \mathbb{R}^n \,:\, \limsup_{y\to x} \frac{u(y)-u(x)-p\cdot(y-x)}{|y-x|} \le 0\right\}.$$

If u is differentiable at x then
$$D^-u(x) = D^+u(x) = \{Du(x)\}.$$

When $u : \mathbb{R}^n \to \mathbb{R}$ is convex, we have the following definition for subgradients. Fix $x_0 \in \mathbb{R}^n$. The subgradient of u at x_0 is defined as
$$\partial u(x_0) = \{p \in \mathbb{R}^n \,:\, u(x) \ge u(x_0) + p\cdot(x-x_0) \text{ for all } x \in \mathbb{R}^n\}.$$

In this convex setting, it is always true that
$$\partial u(x_0) = D^-u(x_0) \neq \emptyset.$$

For $u : \mathbb{R}^n \times [0,\infty) \to \mathbb{R}$ smooth, we write

- $Du(x,t) = D_x u(x,t)$ and $u_t(x,t) = \frac{\partial u}{\partial t}(x,t)$.
- $D^2u(x,t) = D_x^2 u(x,t)$, and $\Delta u(x,t) = \Delta_x u(x,t)$.

Besides, we use the following for a given function $u : \mathbb{R}^n \to \mathbb{R}$.

- Set $u^+ = \max\{u,0\}$, and $u^- = -\min\{u,0\}$. Surely, $u = u^+ - u^-$, and $|u| = u^+ + u^-$.
- If u is compactly supported, then the support of u is denoted by $\mathrm{spt}(u)$.
- If u is $\mathbb{Z}^n$-periodic, then we can think of u as a function from $\mathbb{T}^n$ to $\mathbb{R}$ as well, and vice versa. In the book, we switch freely between the two interpretations.

For a smooth path $\gamma : \mathbb{R} \to \mathbb{R}^n$ and $t \in \mathbb{R}$, we write
$$\dot\gamma(t) = \frac{d}{dt}\gamma(t), \qquad \ddot\gamma(t) = \frac{d^2}{dt^2}\gamma(t).$$

In many occasions, we use a modulus of continuity ω. By this, we mean $\omega : [0,\infty) \to [0,\infty)$ is a continuous function such that $\omega(0) = 0 = \lim_{r\to 0} w(r)$.

The following convolution trick is used quite often throughout the book. Take η to be the standard mollifier, that is,
$$\eta \in C_c^\infty(\mathbb{R}^n, [0,\infty)), \qquad \mathrm{supp}(\eta) \subset B(0,1), \qquad \int_{\mathbb{R}^n} \eta(x)\,dx = 1.$$

For $\varepsilon > 0$, denote by $\eta_\varepsilon(x) = \varepsilon^{-n}\eta\left(\frac{x}{\varepsilon}\right)$ for all $x \in \mathbb{R}^n$. Let $u : \mathbb{R}^n \to \mathbb{R}$ be a continuous function. Set, for $x \in \mathbb{R}^n$,
$$u^\varepsilon(x) = (\eta_\varepsilon \star u)(x) = \int_{\mathbb{R}^n} \eta_\varepsilon(x-y)u(y)\,dy = \int_{B(x,\varepsilon)} \eta_\varepsilon(x-y)u(y)\,dy.$$

Then $u^\varepsilon \in C^\infty(\mathbb{R}^n)$, and $u^\varepsilon \to u$ locally uniformly as $\varepsilon \to 0$. If needed, one can assume further that η is symmetric or radially symmetric.

A.3. Notation for function spaces

- For given $a < b$, $\mathrm{AC}([a,b],\mathbb{R}^n)$ denotes the space of all absolutely continuous curves from $[a,b]$ to $\mathbb{R}^n$. When there is no confusion, we write $\mathrm{AC}([a,b],\mathbb{R}^n)$ as $\mathrm{AC}([a,b])$.
- $C(\mathbb{R}^n) = \{u : \mathbb{R}^n \to \mathbb{R} \,:\, u \text{ is continuous}\}$.
- $B(\mathbb{R}^n) = \{u : \mathbb{R}^n \to \mathbb{R} \,:\, u \text{ is bounded}\}$.
- $\mathrm{BC}(\mathbb{R}^n) = \{u : \mathbb{R}^n \to \mathbb{R} \,:\, u \text{ is bounded, and continuous}\}$.
- $\mathrm{BUC}(\mathbb{R}^n) = \{u \in C(\mathbb{R}^n) \,:\, u \text{ is bounded, uniformly continuous}\}$.
- $C^k(\mathbb{R}^n) = \{u : \mathbb{R}^n \to \mathbb{R} \,:\, u \text{ is } k\text{-times continuously differentiable}\}$, for each given $k \in \mathbb{N}$.
- $C^\infty(\mathbb{R}^n) = \{u : \mathbb{R}^n \to \mathbb{R} \,:\, u \text{ is infinitely differentiable}\}$. For $u \in C^\infty(\mathbb{R}^n)$, we say that u is smooth.
- $C_c^k(\mathbb{R}^n)$, $C_c^\infty(\mathbb{R}^n)$ denote the space of functions in $C^k(\mathbb{R}^n)$, $C^\infty(\mathbb{R}^n)$ that have compact supports, respectively.
- $$\mathrm{Lip}(\mathbb{R}^n) = \big\{u \in C(\mathbb{R}^n) \,:\, \exists\, C > 0 \text{ so that } |u(x) - u(y)| \le C|x-y| \text{ for all } x, y \in \mathbb{R}^n\big\}.$$

  We write
  $$\mathrm{Lip}[u] = \sup_{\substack{x,y\in\mathbb{R}^n\\ x\neq y}} \frac{|u(x)-u(y)|}{|x-y|},$$
  and say that $\mathrm{Lip}[u]$ is the Lipschitz constant of u.
- For $\alpha \in (0,1]$, we say that $u \in C(\mathbb{R}^n)$ is Hölder continuous with exponent α if there exists $C > 0$ such that
  $$|u(x) - u(y)| \le C|x-y|^\alpha \qquad \text{for all } x, y \in \mathbb{R}^n.$$
  In this case, the α-th Hölder seminorm of u is
  $$[u]_{C^{0,\alpha}(\mathbb{R}^n)} = \sup_{\substack{x,y\in\mathbb{R}^n\\ x\neq y}} \frac{|u(x)-u(y)|}{|x-y|^\alpha}.$$
  If we have in addition that u is bounded, then we define the α-th Hölder norm of u to be
  $$\|u\|_{C^{0,\alpha}(\mathbb{R}^n)} = \|u\|_{C(\mathbb{R}^n)} + [u]_{C^{0,\alpha}(\mathbb{R}^n)}.$$
  Then, the Hölder space $C^{0,\alpha}(\mathbb{R}^n)$ is defined as
  $$C^{0,\alpha}(\mathbb{R}^n) = \left\{u \in C(\mathbb{R}^n) \,:\, \|u\|_{C^{0,\alpha}(\mathbb{R}^n)} < +\infty\right\}.$$

- $L^\infty(\mathbb{R}^n) = \big\{u : \mathbb{R}^n \to \mathbb{R} \,:\, u \text{ is Lebesgue measurable, and } \|u\|_{L^\infty(\mathbb{R}^n)} < +\infty\big\}$,
  where
  $$\|u\|_{L^\infty(\mathbb{R}^n)} = \operatorname*{ess\,sup}_{\mathbb{R}^n} |u|.$$
- It is clear that $C^{0,1}(\mathbb{R}^n) = L^\infty(\mathbb{R}^n) \cap \mathrm{Lip}\,(\mathbb{R}^n)$, and $\mathrm{Lip}\,[u] = [u]_{C^{0,1}(\mathbb{R}^n)}$.
- In a same way, one can define $C^{k,\alpha}(\mathbb{R}^n)$ for $k \in \mathbb{N}$ and $\alpha \in (0,1]$.
- $\mathrm{USC}\,(\mathbb{R}^n) = \{u : \mathbb{R}^n \to \mathbb{R} \,:\, u \text{ is upper semicontinuous}\}$.
- $\mathrm{LSC}\,(\mathbb{R}^n) = \{u : \mathbb{R}^n \to \mathbb{R} \,:\, u \text{ is lower semicontinuous}\}$.
- For a function $u : \mathbb{R}^n \to \mathbb{R}$ that is bounded, we denote by
  $$u^*(x) = \limsup_{y \to x} u(y) \qquad \text{for all } x \in \mathbb{R}^n,$$
  and
  $$u_*(x) = \liminf_{y \to x} u(y) \qquad \text{for all } x \in \mathbb{R}^n.$$
  It is clear that $u^* \in \mathrm{USC}\,(\mathbb{R}^n)$, $u_* \in \mathrm{LSC}\,(\mathbb{R}^n)$. We say that u^*, u_* are the upper semicontinuous envelope, and the lower semicontinuous envelope of u, respectively. One has that u is continuous in $\mathbb{R}^n$ if and only if $u^* = u_*$.
- Let $U \subset \mathbb{R}^n$ be a given open set. All above function spaces can be defined in U and $\overline{U}$ in place of $\mathbb{R}^n$ in a similar way.
- $\mathrm{Cvx}\,(\mathbb{R}^n)$ denotes the class of lower semi-continuous convex functions $\phi : \mathbb{R}^n \to \mathbb{R} \cup \{\pm\infty\}$.

A.4. Notation for estimates

- The constants in the estimates are often denoted by C (and C_1, C_2, etc.), which might change from line to line in a given computation. This makes our presentation clearer without keeping track with various factors in each step. Of course, we specify clearly the dependence of these constants on specific parameters.
- (Big-oh notation) For two given functions f, h, we write $f = O(h)$ as $x \to y$ if there exists $C > 0$ such that
  $$|f(x)| \le C|h(x)| \qquad \text{for all } x \text{ sufficiently close to } y.$$
- (Little-oh notation) For two given functions f, h, we write $f = o(h)$ as $x \to y$ if
  $$\lim_{x \to y} \frac{|f(x)|}{|h(x)|} = 0.$$
  In particular, when $h \equiv 1$, we have the notions of $O(1)$ and $o(1)$, respectively.

Appendix B

Some basics on circle homeomorphisms

We give some basics on circle homeomorphisms.

Definition B.1 (Lifted circle homeomorphism)**.** We say that $f : \mathbb{R} \to \mathbb{R}$ is a lifted circle homeomorphism if f is continuous, strictly increasing, and for all $x \in \mathbb{R}$,

$$f(x+1) = f(x) + 1.$$

Sometimes, we simply call f a circle homeomorphism.

If f is a lifted circle homeomorphism, then the Poincaré rotation number

$$\beta_f = \lim_{|i|\to\infty} \frac{f^i(x)}{i}$$

exists and is independent of $x \in \mathbb{R}$. Here, for $i \in \mathbb{N}$, f^i represents the i-th iteration of f. Moreover, for all $i \in \mathbb{Z}$, the periodic function $r_i(x) = f^i(x) - x - i\beta_f$ satisfies

$$|r_i(x)| < 1 \qquad \text{and} \qquad \min_{\mathbb{R}} |r_i| = 0. \tag{B.1}$$

Lemma B.2. *Let f be a lifted circle homeomorphism. Then, $\beta_f = \frac{p}{q} \in \mathbb{Q}$ with $p \in \mathbb{Z}, q \in \mathbb{N}$ if and only if there exists $x_0 \in \mathbb{R}$ such that*

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

Proof. First, if there exists $x_0 \in \mathbb{R}$ such that

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

then by iterations,

$$f^{iq}(x_0) = f(x_0) + ip \qquad \text{for } i \in \mathbb{N}.$$

Hence,

$$\beta_f = \lim_{i\to\infty} \frac{f^{iq}(x_0)}{iq} = \frac{p}{q}.$$

Let us now prove the converse. Assume $\beta_f = \frac{p}{q} \in \mathbb{Q}$. Set

$$g(x) = f^q(x) - x - p \qquad \text{for } x \in \mathbb{R}.$$

Then, g is 1-periodic and continuous. Assume by contradiction that there does not exist $x_0 \in \mathbb{R}$ such that $g(x_0) = 0$. Then either $g > 0$ or $g < 0$. Without loss of generality, assume that

$$\min_{\mathbb{R}} g = \delta > 0.$$

Then, by iterations,

$$f^{iq}(0) \geq ip + i\delta,$$

which yields

$$\liminf_{i\to\infty} \frac{f^{iq}(0)}{iq} \geq \frac{p}{q} + \frac{\delta}{q} > \beta_f,$$

which is absurd. □

Definition B.3. Let f be a lifted circle homeomorphism. We define the set of all recurrent values of f to be

$$\mathrm{Rec}(f) = \overline{\{f^i(x) + k \,:\, i,k \in \mathbb{Z}\}} \subset \mathbb{R}.$$

for any fixed $x \in \mathbb{R}$. Surely, $\mathrm{Rec}(f)$ does not depend on the choice of x.

Proposition B.4. *Let f_1, f_2 be two lifted circle homeomorphisms. Assume that $\beta_{f_1} = \beta_{f_2} = \beta \in \mathbb{R} \setminus \mathbb{T}^n \times (0,\infty)$. Then either $\mathrm{Rec}(f_1) = \mathrm{Rec}(f_2)$ and $f_1|_{\mathrm{Rec}(f_1)} = f_2|_{\mathrm{Rec}(f_2)}$ or there exist $x_1 \in \mathrm{Rec}(f_1)$ and $x_2 \in \mathrm{Rec}(f_2)$ such that the orbits $(f_1^i(x_1))_{i\in\mathbb{Z}}$ and $(f_2^i(x_2))_{i\in\mathbb{Z}}$ cross infinitely often.*

Proof. As β is irrational, we have one basic but important point that for any $x_0 \in \mathbb{R}$ and $l = 1, 2$,

$$j\beta + k \mapsto f_l^j(x_0) + k \text{ is strictly increasing.}$$

Denote by, for $l = 1, 2$,

$$x_l^+(t) = \inf\left\{ f_l^j(x_0) + k \,:\, j\beta + k > t \right\},$$

$$x_l^-(t) = \sup\left\{ f_l^j(x_0) + k \,:\, j\beta + k < t \right\}.$$

We have the following basic properties of $x_l^\pm$.

(1) $x_l^\pm$ are strictly increasing for $l = 1, 2$.

(2) x_l^+ is continuous from the right; and x_l^- is continuous from the left.

(3) For each $l = 1, 2$ fixed, x_l^+ and x_l^- are continuous at the same points and coincide at such points.
(4) $x_l^{\pm}(t+1) = x_l^{\pm}(t) + 1$.
(5) $f_l \circ x_l^{\pm}(t) = x_l^{\pm}(t+\beta)$.
(6) $\mathrm{Rec}(f_l) = x_l^+(\mathbb{R}) \cup x_l^-(\mathbb{R})$.

Thanks to (1) and (5), for each $l = 1, 2$ fixed, if $x_l^{\pm}$ are not continuous, then they have upward jumps on a countable dense set of $\mathbb{R}$. On the other hand, if $x_l^+ = x_l^- = x_l$ is continuous, then x_l is a lifted circle homeomorphism, and in such case,

$$x_l^- \circ f_l \circ x_l(t) = t + \beta.$$

Let us now proceed to prove the claims. There are two cases to be considered.

Case 1. There exists $c \in \mathbb{R}$ such that $x_1^-(\cdot + c) - x_2^-(\cdot)$ changes sign. Then, we can find nonempty open intervals I_1, I_2 such that

$$x_1^-(t+c) < x_2^-(t) \qquad \text{for } x \in I_1 + \mathbb{Z},$$
$$x_1^-(t+c) > x_2^-(t) \qquad \text{for } x \in I_2 + \mathbb{Z}.$$

Set

$$x_1 = x_1^-(c) \in \mathrm{Rec}(f_1), \qquad x_2 = x_2^-(0) \in \mathrm{Rec}(f_2).$$

Then, for $j \in \mathbb{Z}$,

$$f_1^j(x_1) = x_1^-(c + j\beta), \qquad f_2^j(x_2) = x_2^-(j\beta).$$

As β is irrational, $\{j\beta\}_{j\in\mathbb{N}}$ is in each of $I_1 + Z$ and $I_2 + Z$ infinitely often. Thus, the orbits $(f_1^i(x_1))_{i\in\mathbb{Z}}$ and $(f_2^i(x_2))_{i\in\mathbb{Z}}$ cross infinitely often.

Case 2. For every $c \in \mathbb{R}$, $x_1^-(\cdot + c) - x_2^-(\cdot)$ does not change sign. Set

$$c_0 = \sup \left\{ c \in \mathbb{R} \,:\, x_1^-(\cdot + c) \le x_2^-(\cdot) \right\}.$$

Because of (4), c_0 exists and is finite. Thanks to (3), we have $x_1^-(\cdot + c_0) \le x_2^-(\cdot)$. If there exists $t_0 \in \mathbb{R}$ such that x_1^- is continuous at $t_0 + c_0$ and

$$x_1^-(t_0 + c_0) < x_2^-(t_0),$$

then there exists $c_1 > c_0$ such that

$$x_1^-(t_0 + c_1) < x_2^-(t_0).$$

This contradicts the definition of c_0 and the situation in Case 2. Thus, if x_1^- is continuous at $t + c_0$ for $t \in \mathbb{R}$, then

$$x_1^-(t + c_0) = x_2^-(t).$$

By the denseness of continuous points of x_1^-, we yield

$$x_1^{\pm}(t + c_0) = x_2^{\pm}(t) \qquad \text{for } t \in \mathbb{R}.$$

Hence, $\mathrm{Rec}(f_1) = \mathrm{Rec}(f_2)$, and

$$f_1(x_1^{\pm}(t + c_0)) = x_1^{\pm}(t + c_0 + \beta) = x_2^{\pm}(t + \beta) = f_2(x_2^{\pm}(t)).$$

We conclude that $f_1|_{\mathrm{Rec}(f_1)} = f_2|_{\mathrm{Rec}(f_2)}$. $\square$

Appendix C

The method of characteristics for Hamilton–Jacobi equations

C.1. Quick overview

We give a brief overview of the method of characteristics for Hamilton–Jacobi equations. Our main object here is

$$\begin{cases} u_t + H(x, Du) = 0 & \text{in } \mathbb{R}^n \times \mathbb{R}, \\ u(x,0) = g(x) & \text{on } \mathbb{R}^n. \end{cases} \tag{C.1}$$

Here, the Hamiltonian $H \in C^2(\mathbb{R}^n \times \mathbb{R}^n)$ is given. We assume that the initial data $g \in C^2(\mathbb{R}^n)$ and $\|g\|_{C^2(\mathbb{R}^n)} < +\infty$.

We aim at solving (C.1) locally in time by converting the PDE into an appropriate system of ODE. For $(x,t) \in \mathbb{R}^n \times \mathbb{R}$, we would like to calculate $u(x,t)$ by find a curve in $\mathbb{R}^n \times \mathbb{R}$ connecting (x,t) with some $(x_0, 0) \in \mathbb{R}^n \times \mathbb{R}$. Since we know that $u(x_0, 0) = g(x_0)$, we hope to be able to calculate u along this particular curve, and hence obtain $u(x,t)$. Let us write this curve as $\mathbf{x}(t)$ with $\mathbf{x}(0) = x_0$. For $t \in \mathbb{R}$, set

$$\begin{cases} \mathbf{p}(t) = Du(\mathbf{x}(t), t), \\ z(t) = u(\mathbf{x}(t), t). \end{cases} \tag{C.2}$$

Assume for now that everything is smooth and nice. For $1 \leq i \leq n$, as $p_i(t) = u_{x_i}(\mathbf{x}(t), t)$,

$$\dot{p}_i(t) = u_{x_i t}(\mathbf{x}(t), t) + Du_{x_i}(\mathbf{x}(t), t) \cdot \dot{\mathbf{x}}(t).$$

On the other hand, differentiate eqrefeq:HJ-char with respect to x_i to yield

$$u_{x_i t} + D_p H(x, Du) \cdot Du_{x_i} + H_{x_i}(x, Du) = 0.$$

From the two relations above, it is quite natural to choose $\mathbf{x}(t)$ such that

$$\dot{\mathbf{x}}(t) = D_p H(\mathbf{x}(t), Du(\mathbf{x}(t))) = D_p H(\mathbf{x}(t), \mathbf{p}(t)).$$

Then, we also get

$$\dot{\mathbf{p}}(t) = -D_x H(\mathbf{x}(t), \mathbf{p}(t)).$$

Combining the two, we arrive at exactly the Hamiltonian system

$$\begin{cases} \dot{\mathbf{x}}(t) = D_p H(\mathbf{x}(t), \mathbf{p}(t)), \\ \dot{\mathbf{p}}(t) = -D_x H(\mathbf{x}(t), \mathbf{p}(t)). \end{cases} \tag{C.3}$$

Moreover, as $z(t) = u(\mathbf{x}(t), t)$,

$$\dot{z}(t) = u_t(\mathbf{x}(t), t) + Du(\mathbf{x}(t), t) \cdot \dot{\mathbf{x}}(t) = -H(\mathbf{x}(t), \mathbf{p}(t)) + \mathbf{p}(t) \cdot D_p H(\mathbf{x}(t), \mathbf{p}(t)).$$

C.2. Method of characteristics

For the ODE system of characteristics, we also include the equation for z in the Hamiltonian system, that is,

$$\begin{cases} \dot{\mathbf{x}}(t) = D_p H(\mathbf{x}(t), \mathbf{p}(t)), \\ \dot{\mathbf{p}}(t) = -D_x H(\mathbf{x}(t), \mathbf{p}(t)), \\ \dot{z}(t) = \mathbf{p}(t) \cdot D_p H(\mathbf{x}(t), \mathbf{p}(t)) - H(\mathbf{x}(t), \mathbf{p}(t)). \end{cases} \tag{C.4}$$

Here, the corresponding initial data is, for given $y \in \mathbb{R}^n$,

$$\begin{cases} \mathbf{x}(0) = y, \\ \mathbf{p}(0) = Dg(y), \\ z(0) = g(y). \end{cases}$$

To demonstrate clearly the dependence on initial data, we write

$$\begin{cases} \mathbf{x}(t) = \mathbf{x}(y, t), \\ \mathbf{p}(t) = \mathbf{p}(y, t), \\ z(t) = z(y, t). \end{cases}$$

Lemma C.1 (Local invertibility)**.** *Fix $x_0 \in \mathbb{R}^n$. There exist an open interval $I \subset \mathbb{R}$ containing 0, and two neighborhoods V, W of x_0 in $\mathbb{R}^n$ such that, for each $(x, t) \in V \times I$, there exists a unique $y \in W$ such that*

$$x = \mathbf{x}(y, t).$$

Moreover, the map $(x, t) \mapsto y$ is C^2.

Proof. Consider the map $(y,t) \mapsto G(y,t) = (\mathbf{x}(y,t),t)$. At $(y,0)$, we see that

$$DG(y,0) = \begin{bmatrix} 1 & 0 & \cdots & 0 & H_{p_1} \\ 0 & 1 & \cdots & 0 & H_{p_2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & H_{p_n} \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

Here DG is the gradient of G in (y,t) variable, and is a square matrix of size $n+1$. And H_{p_i} is being evaluated at $(y, Dg(y))$ for $1 \le i \le n$. Of course

$$\det DG(y,0) = 1.$$

By the inverse function theorem, we get the desired conclusion. $\square$

In view of the above lemma, for each $(x,t) \in V \times I$, we can locally uniquely solve the equation

$$x = \mathbf{x}(y,t) \qquad \text{for } y = \mathbf{y}(x,t) \in C^2.$$

Denote by

$$\text{(C.5)} \qquad \begin{cases} \mathbf{p}(x,t) = \mathbf{p}(\mathbf{y}(x,t),t), \\ u(x,t) = z(\mathbf{y}(x,t),t). \end{cases}$$

Theorem C.2 (Local existence theorem). *The function u defined in* (C.5) *above is in $C^2(V \times I)$ and solves* (C.1) *in $V \times I$ with initial condition*

$$u(x,0) = g(x) \qquad \textit{for } x \in \mathbb{R}^n.$$

Proof. $\square$

Proof. By (C.4), we first have

$$\text{(C.6)} \qquad z_t(y,t) = \mathbf{p}(y,t) \cdot \mathbf{x}_t(y,t) - H(\mathbf{x}(y,t), \mathbf{p}(y,t)).$$

We claim that, for $1 \le i \le n$,

$$\text{(C.7)} \qquad z_{y_i}(y,t) = \mathbf{p}(y,t) \cdot \mathbf{x}_{y_i}(y,t).$$

Indeed, denote by

$$r^i(t) = z_{y_i}(y,t) - \mathbf{p}(y,t) \cdot \mathbf{x}_{y_i}(y,t) \qquad \text{for } t \in \mathbb{R}.$$

Clearly, $r^i(0) = 0$, and

$$\dot{r}^i(t) = z_{y_i t}(y,t) - \mathbf{p}_t(y,t) \cdot \mathbf{x}_{y_i}(y,t) - \mathbf{p}(y,t) \cdot \mathbf{x}_{y_i t}(y,t).$$

Differentiate (C.6) with respect to y_i and use (C.4) to yield

$$\begin{aligned} z_{y_i t}(y,t) &= \mathbf{p} \cdot \mathbf{x}_{y_i t} + \mathbf{p}_{y_i} \cdot \mathbf{x}_t - D_x H \cdot \mathbf{x}_{y_i} - D_p H \cdot \mathbf{p}_{y_i} \\ &= \mathbf{p} \cdot \mathbf{x}_{y_i t} + \mathbf{p}_t \cdot \mathbf{x}_{y_i}. \end{aligned}$$

Combine this with the relation above to yield $\dot{r}^i(t) = 0$, and hence $r^i \equiv 0$. We get (C.7).

Next, we show that

$$\mathbf{p}(x,t) = Du(x,t). \tag{C.8}$$

Indeed, for $1 \le k \le n$, thanks to (C.7),

$$\begin{aligned} u_{x_k}(x,t) &= z_{y_i}(\mathbf{y}(x,t),t)(y_i)_{x_k} \\ &= \mathbf{p}(\mathbf{y}(x,t),t) \cdot \mathbf{x}_{y_i}(\mathbf{y}(x,t),t)(y_i)_{x_k} \\ &= \mathbf{p}(\mathbf{y}(x,t),t) \cdot x_{x_k} = p_k(x,t), \end{aligned}$$

which gives (C.8).

Finally, we prove that u defined in (C.5) solves (C.1) in $V \times I$. Note that $\mathbf{x}(\mathbf{y}(x,t),t) = x$, which gives

$$\mathbf{x}_t + \mathbf{x}_{y_i}(y_i)_t = 0.$$

We use this, (C.5), and (C.6)–(C.8) to compute

$$\begin{aligned} u_t(x,t) &= z_t + z_{y_i}(y_i)_t \\ &= \mathbf{p} \cdot \mathbf{x}_t - H(x, Du(x,t)) + \mathbf{p} \cdot \mathbf{x}_{y_i}(y_i)_t \\ &= -H(x, Du(x,t)) + \mathbf{p} \cdot (\mathbf{x}_t + \mathbf{x}_{y_i}(y_i)_t) = -H(x, Du(x,t)). \end{aligned}$$

The proof is complete. □

We note that the above theorem is only about a local existence result defined in $V \times I \subset \mathbb{R}^n \times \mathbb{R}$, a neighborhood of $(x_0, 0) \in \mathbb{R}^n \times \mathbb{R}$. This solution $u \in C^2(V \times I)$ found by the method of characteristics is the same as the viscosity solution to (C.1) in $V \times I$ provided that the Hamiltonian H satisfies some appropriate structural assumptions. A common structural assumption that we put on H is

$$\begin{cases} H \in \mathrm{BUC}\,(\mathbb{R}^n \times B(0,R)) & \text{for all } R > 0, \\ \lim_{|p| \to \infty} \inf_{x \in \mathbb{R}^n} H(x,p) = +\infty. \end{cases}$$

Under this assumption and the condition that $g \in C^2(\mathbb{R}^n)$ with $\|g\|_{C^2(\mathbb{R}^n)} < +\infty$, (C.1) has a unique viscosity solution $u \in \mathrm{Lip}\,(\mathbb{R}^n \times [0,\infty))$. Assume a bit more that $H \in \mathrm{Lip}\,(\mathbb{R}^n \times B(0,R))$ for all $R > 0$. Then, by [**Tra21**, Theorem 1.39], u has the finite speed of propagation property. This can be seen directly from the method of characteristics too. Set $\|Du\|_{L^\infty(\mathbb{R}^n \times [0,\infty))} = R > 0$, and $\|D_x H\|_{L^\infty(\mathbb{R}^n \times B(0,R))} + \|D_p H\|_{L^\infty(\mathbb{R}^n \times B(0,R))} = C > 0$. Then, $|\mathbf{p}(t)| \le R$, and thanks to (C.3),

$$|\dot{\mathbf{p}}(t)| + |\dot{\mathbf{x}}(t)| \le C.$$

In particular, the values of $u(x,t)$ for $(x,t) \in \overline{B}(0,r) \times [0,T]$ for $r, T > 0$ are determined by the values of g on $\overline{B}(0, r + CT)$. Hence, we get that the local existence theorem is consistent with the finite speed of propagation property.

C.3. References

(1) For the method of characteristics for general first-order equations, we refer the readers to Chapter 3 in Evans [**Eva10**]. We here only give the method of characteristics for Hamilton-Jacobi equations for our purposes.

(2) For the finite speed of propagation of solutions to Hamilton-Jacobi equations, see [**Tra21**, Theorem 1.39].

Appendix D

The Fekete lemma

D.1. Fekete's lemma

Here is the main result.

Lemma D.1 (Fekete's lemma). *Let $\phi : (0, \infty) \to \mathbb{R}$ be measurable and subadditive, that is, for $l, r > 0$,*

$$\phi(l + r) \leq \phi(l) + \phi(r).$$

Then,

$$\lim_{k\to\infty} \frac{\phi(k)}{k} = \inf_{l>0} \frac{\phi(l)}{l}.$$

We first prove that ϕ is bounded on $(0, a]$ for any given $a > 0$.

Lemma D.2. *Let $\phi : (0, \infty) \to \mathbb{R}$ be measurable and subadditive. Then, for any given $a > 0$, $\sup_{(0,a]} \phi < +\infty$.*

Proof. Assume by contradiction that there exists a sequence $\{t_n\} \subset (0, a]$ such that

$$\phi(t_n) > 2n.$$

As ϕ is subadditive, for $s \in (0, t_n)$,

$$\phi(s) + \phi(t_n - s) \geq \phi(t_n) > 2n \quad \Rightarrow \quad \max\{\phi(s), \phi(t_n - s)\} > n.$$

For $n \in \mathbb{N}$, denote by $E_n = \phi^{-1}([n, \infty)) \cap (0, a]$. Then, thanks to the above, $|E_n| \geq a/2$. As $\{E_n\}$ is a nested sequence of sets, we yield that

$$\left|\phi^{-1}(\{+\infty\}) \cap (0, a]\right| = \left|\bigcap_{n\in\mathbb{N}} E_n\right| \geq a/2,$$

which gives a contradiction. $\square$

We are now ready to prove Fekete's lemma.

Proof of Lemma D.1. Let

$$L = \inf_{l>0} \frac{\phi(l)}{l} \in [-\infty, \infty).$$

Note that L could be $-\infty$. It is obvious from the definition of L that, for $k > 0$,

$$\frac{\phi(k)}{k} \geq L.$$

Fix $c > L$. There exists $l_0 > 0$ such that

$$\frac{\phi(l_0)}{l_0} \leq c.$$

Thanks to Lemma D.2,

$$\sup_{(0,l_0]} \phi = M < +\infty.$$

For $nl_0 < k \leq (n+1)l_0$ for $n \in \mathbb{N}$, we use the above and the subadditivity of ϕ to imply

$$\phi(k) \leq \phi(nl_0) + \phi(k - nl_0) \leq n\phi(l_0) + M,$$

and hence

$$\frac{\phi(k)}{k} \leq \frac{n\phi(l_0)}{nl_0} + \frac{C}{nl_0} \leq c + \frac{M}{nl_0}.$$

Therefore,

$$\limsup_{k\to\infty} \frac{\phi(k)}{k} \leq c.$$

The proof is complete. □

We note that the proof of Fekete's lemma is rather elementary and simple. The proof is qualitative and, in general, there is no convergence rate of $\phi(k)/k$ to L.

D.2. References

(1) For more discussions on subadditive functions, we refer the readers to Hille, Phillips [**HP57**, Chapter 7].

Bibliography

[AAM09] Shiri Artstein-Avidan and Vitali Milman, *The concept of duality in convex analysis, and the characterization of the Legendre transform*, Ann. of Math. (2) **169** (2009), no. 2, 661–674. MR 2480615

[ABIL13] Yves Achdou, Guy Barles, Hitoshi Ishii, and Grigory L. Litvinov, *Hamilton-Jacobi equations: approximations, numerical analysis and applications*, Lecture Notes in Mathematics, vol. 2074, Springer, Heidelberg; Fondazione C.I.M.E., Florence, 2013, Lecture Notes from the CIME Summer School held in Cetraro, August 29–September 3, 2011, Edited by Paola Loreti and Nicoletta Anna Tchou, Fondazione CIME/CIME Foundation Subseries. MR 3135343

[Ale90] Kenneth S. Alexander, *Lower bounds on the connectivity function in all directions for Bernoulli percolation in two and three dimensions*, Ann. Probab. **18** (1990), no. 4, 1547–1562. MR 1071808

[Ale97] ______, *Approximation of subadditive functions and convergence rates in limiting-shape results*, Ann. Probab. **25** (1997), no. 1, 30–55. MR 1428498

[Ban88] V. Bangert, *Mather sets for twist maps and geodesics on tori*, Dynamics reported, Vol. 1, Dynam. Report. Ser. Dynam. Systems Appl., vol. 1, Wiley, Chichester, 1988, pp. 1–56. MR 945963

[Ban94] Victor Bangert, *Geodesic rays, Busemann functions and monotone twist maps*, Calc. Var. Partial Differential Equations **2** (1994), no. 1, 49–63. MR 1384394

[BS00a] G. Barles and Panagiotis E. Souganidis, *On the large time behavior of solutions of Hamilton-Jacobi equations*, SIAM J. Math. Anal. **31** (2000), no. 4, 925–939. MR 1752423

[BS00b] Guy Barles and Panagiotis E. Souganidis, *Some counterexamples on the asymptotic behavior of the solutions of Hamilton-Jacobi equations*, C. R. Acad. Sci. Paris Sér. I Math. **330** (2000), no. 11, 963–968. MR 1779687

[BS01] G. Barles and P. E. Souganidis, *Space-time periodic solutions and long-time behavior of solutions to quasi-linear parabolic equations*, SIAM J. Math. Anal. **32** (2001), no. 6, 1311–1323. MR 1856250

[Bur92] D. Yu. Burago, *Periodic metrics*, Representation theory and dynamical systems, Adv. Soviet Math., vol. 9, Amer. Math. Soc., Providence, RI, 1992, pp. 205–210. MR 1166203

[Car95] M. J. Dias Carneiro, *On minimizing measures of the action of autonomous Lagrangians*, Nonlinearity **8** (1995), no. 6, 1077–1085. MR 1363400

[CDI01] I. Capuzzo-Dolcetta and H. Ishii, *On the rate of convergence in homogenization of Hamilton-Jacobi equations*, Indiana Univ. Math. J. **50** (2001), no. 3, 1113–1129. MR 1871349

[CGMT15] Filippo Cagnetti, Diogo Gomes, Hiroyoshi Mitake, and Hung V. Tran, *A new method for large time behavior of degenerate viscous Hamilton-Jacobi equations with convex Hamiltonians*, Ann. Inst. H. Poincaré Anal. Non Linéaire **32** (2015), no. 1, 183–200. MR 3303946

[Coo22] William Cooperman, *A near-optimal rate of periodic homogenization for convex Hamilton-Jacobi equations*, Arch. Ration. Mech. Anal. **245** (2022), no. 2, 809–817. MR 4451475

[CS04] Piermarco Cannarsa and Carlo Sinestrari, *Semiconcave functions, Hamilton-Jacobi equations, and optimal control*, Progress in Nonlinear Differential Equations and their Applications, vol. 58, Birkhäuser Boston, Inc., Boston, MA, 2004. MR 2041617

[DS06] Andrea Davini and Antonio Siconolfi, *A generalized dynamical approach to the large time behavior of solutions of Hamilton-Jacobi equations*, SIAM J. Math. Anal. **38** (2006), no. 2, 478–502. MR 2237158

[Eva04] Lawrence C. Evans, *A survey of partial differential equations methods in weak KAM theory*, Comm. Pure Appl. Math. **57** (2004), no. 4, 445–480. MR 2026176

[Eva08] ———, *Weak KAM theory and partial differential equations*, Calculus of variations and nonlinear partial differential equations, Lecture Notes in Math., vol. 1927, Springer, Berlin, 2008, pp. 123–154. MR 2408260

[Eva10] ———, *Partial differential equations*, second ed., Graduate Studies in Mathematics, vol. 19, American Mathematical Society, Providence, RI, 2010. MR 2597943

[Fat] Albert Fathi, *Weak kam theorem in lagrangian dynamics*, to appear in Cambridge Studies in Advanced Mathematics.

[FS04] Albert Fathi and Antonio Siconolfi, *Existence of C^1 critical subsolutions of the Hamilton-Jacobi equation*, Invent. Math. **155** (2004), no. 2, 363–388. MR 2031431

[FS05] ———, *PDE aspects of Aubry-Mather theory for quasiconvex Hamiltonians*, Calc. Var. Partial Differential Equations **22** (2005), no. 2, 185–228. MR 2106767

[GMT21] Diogo A. Gomes, Hiroyoshi Mitake, and Hung V. Tran, *The large time profile for Hamilton–Jacobi–Bellman equations*, Mathematische Annalen (2021).

[Gom09] Diogo Gomes, *Viscosity solutions of Hamilton-Jacobi equations*, Publicações Matemáticas do IMPA. [IMPA Mathematical Publications], Instituto Nacional de Matemática Pura e Aplicada (IMPA), Rio de Janeiro, 2009, 27o Colóquio Brasileiro de Matemática. [27th Brazilian Mathematics Colloquium]. MR 2536761

[HP57] Einar Hille and Ralph S. Phillips, *Functional analysis and semi-groups*, vol. Vol. 31, American Mathematical Society, Providence, R.I., 1957, rev. ed. MR 89373

[Ish] Hitoshi Ishii, *Lecture notes on the weak kam theorem*, lecture notes available on Ishii's website.

[Ish08] ———, *Asymptotic solutions for large time of Hamilton-Jacobi equations in Euclidean n space*, Ann. Inst. H. Poincaré Anal. Non Linéaire **25** (2008), no. 2, 231–266. MR 2396521

[JTY20] Wenjia Jing, Hung V. Tran, and Yifeng Yu, *Effective fronts of polytope shapes*, Minimax Theory Appl. **5** (2020), no. 2, 347–360. MR 4132072

[JTY21] ———, *Effective fronts of polygon shapes in two dimensions*, 2021.

[Kal05] Vadim Yu. Kaloshin, *Mather theory, weak KAM theory, and viscosity solutions of Hamilton-Jacobi PDE's*, EQUADIFF 2003, World Sci. Publ., Hackensack, NJ, 2005, pp. 39–48. MR 2185002

[LMT17] Nam Q. Le, Hiroyoshi Mitake, and Hung V. Tran, *Dynamical and geometric aspects of Hamilton-Jacobi and linearized Monge-Ampère equations—VIASM 2016*, Lecture Notes in Mathematics, vol. 2183, Springer, Cham, 2017, Edited by Mitake and Tran. MR 3729436

[LPV] Pierre-Louis Lions, George Papanicolaou, and S. R. Srinivasa Varadhan, *Homogenization of hamilton–jacobi equations*, unpublished work (1987).

[MTY19] Hiroyoshi Mitake, Hung V. Tran, and Yifeng Yu, *Rate of convergence in periodic homogenization of Hamilton-Jacobi equations: the convex setting*, Arch. Ration. Mech. Anal. **233** (2019), no. 2, 901–934. MR 3951696

[NR99] Gawtum Namah and Jean-Michel Roquejoffre, *Remarks on the long time behaviour of the solutions of Hamilton-Jacobi equations*, Comm. Partial Differential Equations **24** (1999), no. 5-6, 883–893. MR 1680905

[Sor15] Alfonso Sorrentino, *Action-minimizing methods in Hamiltonian dynamics*, Mathematical Notes, vol. 50, Princeton University Press, Princeton, NJ, 2015, An introduction to Aubry-Mather theory. MR 3330134

[Tra21] H.V. Tran, *Hamilton–Jacobi equations: Theory and applications*, Graduate Studies in Mathematics, American Mathematical Society, 2021.

[Tu] Son N. T. Tu, *Rate of convergence for periodic homogenization of convex hamilton-jacobi equations in one dimension*, Asymptotic Analysis, vol. Pre-press, no. Pre-press, pp. 1–24, 2020.

[TY21] Hung V. Tran and Yifeng Yu, *Optimal convergence rate for periodic homogenization of convex Hamilton-Jacobi equations*, 2021.

[TY22] ———, *Differentiability of effective fronts in the continuous setting in two dimensions*, 2022.

Index