

UC Berkeley

UC Berkeley Previously Published Works

Title

Continuity of optimal control costs and its application to weak KAM theory

Permalink

<https://escholarship.org/uc/item/18b5m8kp>

Journal

Calculus of Variations and Partial Differential Equations, 39(1)

ISSN

1432-0835

Authors

Agrachev, Andrei
Lee, Paul W.

Publication Date

2010-09-01

DOI

10.1007/s00526-010-0308-4

Peer reviewed

Continuity of optimal control costs and its application to weak KAM theory

Andrei Agrachev · Paul W. Y. Lee

Received: 23 September 2009 / Accepted: 13 January 2010 / Published online: 3 February 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract We prove continuity of certain cost functions arising from optimal control of affine control systems. We give sharp sufficient conditions for this continuity. As an application, we prove a version of weak KAM theorem and consider the Aubry–Mather problems corresponding to these systems.

Mathematics Subject Classification (2000) 35F21 · 49L25

1 Introduction

Integrability of Hamiltonian systems has been a subject of considerable interest for several decades. One way to understand the dynamics of such systems is to find a family of smooth solutions, called generating functions, to the time-independent Hamilton–Jacobi equation. These generating functions define symplectic transformations which transform the given completely integrable Hamiltonian system to a much simpler one that are easily solvable.

On the contrary, if the Hamiltonian system is not completely integrable, then it is natural to ask whether one can solve the Hamilton–Jacobi equation in certain weak sense. This is accomplished in, what is known as, the weak KAM theorem under certain assumptions on

Communicated by L. Ambrosio.

A. Agrachev (✉)
International School for Advanced Studies, via Beirut 4, 34014 Trieste, Italy
e-mail: agrachev@sissa.it

A. Agrachev
Steklov Mathematical Institute, ul. Gubkina 8, Moscow 119991, Russia

P. W. Y. Lee
Department of Mathematics, University of California at Berkeley, 970 Evans Hall #3840,
Berkeley, CA 94720-3840, USA
e-mail: plee@math.berkeley.edu

the Hamiltonian. More precisely, let $L : TM \rightarrow \mathbb{R}$ be a Lagrangian defined on the tangent bundle TM of a compact manifold M which satisfies the following conditions:

- (1) the restriction of the Lagrangian L to each tangent space has positive definite Hessian,
- (2) $L(x, v) \geq C|v|^2 + K$ for some Riemannian metric $|\cdot|$ and some constants $K, C > 0$.

Let $H : T^*M \rightarrow \mathbb{R}$ be the corresponding Hamiltonian defined by the Legendre transform:

$$H(x, \alpha) = \sup_{v \in T_x M} [\alpha(v) - L(x, v)].$$

The following is the weak KAM theorem mentioned above. It was first proven in [12] when M is a torus and was extended to all compact manifolds in [8] (see also [10] for a version related to vakonomic mechanics).

Theorem 1.1 *Under the above assumptions, there exists a unique constant h such that the Hamilton–Jacobi equation*

$$H(x, df_x) = -h, \tag{1.1}$$

has a viscosity solution.

In order to give the definition of viscosity solution, we first recall the concepts of sub- and super-differentials. If f is a continuous function on a manifold M , then the sub-differential $d^- f_x$ of the function f at a point x is the subset of the cotangent space T_x^*M defined by the following: a co-vector p in the cotangent space T_x^*M is contained in the sub-differential $d^- f_x$ of f at x if there exists a smooth function g defined in a neighborhood O of x such that $dg_x = p$ and g touches f from above. By g touching from above, we mean that $f(x) = g(x)$ and $f(y) \leq g(y)$ for all y in the set O . The super-differential $d^+ f$ of f is defined in a similar way with the function g touching from below instead. Let $G : \mathbb{R} \times T^*M \rightarrow \mathbb{R}$ be a continuous function, then a continuous function f is called a sub-solution to the equation $G(f(x), x, p) = 0$ if for each p in the sub-differential $d^- f_x$,

$$G(f(x), x, p) \leq 0.$$

Similarly, f is a super-solution if for each p in the super-differential $d^+ f_x$,

$$G(f(x), x, p) \geq 0.$$

If f is both a super and a sub-solution, then it is called a viscosity solution (see [6] for various different characterizations of the sub-differential and viscosity solution).

In this paper, we study weak KAM theorem corresponding to Hamiltonians which arise from certain optimal control problems. More precisely, let X_0, X_1, \dots, X_n be smooth vector fields on a compact manifold M of dimension m and consider the following family of ODEs, called control-affine system:

$$\dot{x}(t) = F(x(t), u(t)) := X_0(x(t)) + \sum_{i=1}^n u_i(t)X_i(x(t)), \tag{1.2}$$

where $u(\cdot) := (u_1(\cdot), \dots, u_n(\cdot)) : [0, T] \rightarrow \mathbb{R}^n$ are essentially bounded measurable functions, called controls, and solutions to (1.2) are Lipschitz curves in M , called admissible paths.

Let $L : M \times \mathbb{R}^n \rightarrow \mathbb{R}$ be a smooth function, called Lagrangian. The optimal control cost c_T corresponding to the above control affine system (1.2) and Lagrangian L is the following function:

$$c_T(x, y) = \inf \int_0^T L(x(t), u(t))dt, \tag{1.3}$$

where the infimum is taken over all pairs $(x(\cdot), u(\cdot))$ which satisfies the affine control system (1.2) and the boundary conditions $x(0) = x$ and $x(T) = y$.

Since there may exist points which are not connected by any admissible path, the above cost function is not always well-defined without additional assumptions. We recall that a family of vector fields $\{X_1, \dots, X_n\}$ is said to be k -generating if the vector fields X_i and their iterated Lie brackets up to $k - 1$ order spanned each tangent space in TM . More precisely, the following holds for each point x in the manifold M

$$T_x M = \text{span}\{[X_{i_1}, [X_{i_2}, \dots, [X_{i_{l-1}}, X_{i_l}]](x) \mid 1 \leq i_j \leq n, 1 \leq l \leq k\}.$$

The family $\{X_1, \dots, X_n\}$ is bracket generating if it is k -generating for some k . If we assume that the family $\{X_1, \dots, X_n\}$ is bracket generating, then any two points can be connected by an admissible path [4]. Therefore, under this assumption, the cost c_T in (1.3) is well-defined for any $T > 0$ and any points x, y on the manifold M .

In this paper, we prove continuity of the optimal control cost c_T under some growth and convexity conditions on the Lagrangian L (see Theorem 3.2). A simple useful corollary of the general continuity result is as follows:

Theorem 1.2 *Assume that the Lagrangian L and the vector fields X_1, \dots, X_n satisfy the following conditions:*

- (1) $C_1|u|^q + K_1 \leq L(x, u) \leq C_2|u|^2 + K_2,$
- (2) $\left| \frac{\partial L(x, u)}{\partial x} \right| \leq C_3|u|^2,$
- (3) *the Hessian of L in the u variable is positive definite, and*
- (4) $\{X_1, \dots, X_n\}$ *is 3-generating*

for some constants $C_1, C_2, C_3, K_1, K_2 > 0$ and some constant $q > 1$. Then the cost function $(t, x, y) \mapsto c_t(x, y)$ defined in (1.3) is continuous.

As an application, we prove a version of the weak KAM theorem corresponding to the above optimal control cost c . More precisely, let $H : T^*M \rightarrow \mathbb{R}$ be the Hamiltonian function defined by

$$H(x, \alpha_x) = \sup_{u \in U} [\alpha_x(F(x, u)) - L(x, u)] \tag{1.4}$$

Note that the Hamiltonian H is, in general, neither fiberwise strictly convex nor coercive, which are basic assumptions on the classical weak KAM theory (see [9]).

Theorem 1.3 (Weak KAM Theorem) *If we make the same assumptions as in Theorem 1.2, then there exists a unique constant h such that the Hamilton–Jacobi equation (1.1) has a viscosity solution.*

The structure of this paper is as follows. In Sect. 2, we give a counter example showing that the 3-generating condition in Theorem 1.2 is essential. Sections 3 and 4 are devoted to the proof of Theorems 1.2 and 1.3, respectively. In Sect. 5, we study a generalization of the Aubry–Mather problem to the present setting.

2 Example

Assume that M is two-dimensional and the control system has the form:

$$\dot{x}_1 = u_1, \quad \dot{x}_2 = x_1^2 + u_2 x_1^k,$$

in some local coordinate chart. The family of vector fields $X_1(x_1, x_2) = (1, 0)$ and $X_2(x_1, x_2) = (0, x_1^k)$ is $(k + 1)$ -generating but not k -generating. In this section, we show that the cost function c_1 corresponding to the Lagrangian $L(x, u) = \frac{1}{2} (u_1^2 + u_2^2)$ is not continuous if $k \geq 3$. This shows that the 3-generating assumption in Theorem 1.2 is essential. More precisely,

Proposition 2.1 *Assume that $k \geq 3$. Then the cost function c_1 corresponding to the above control system and Lagrangian satisfies*

$$c_1((0, w), (0, w)) = 0, \quad c_1((0, w), (0, z)) \geq K$$

for some constant $K > 0$, all w , and all $z < 0$. In particular, the cost function c_1 is not continuous.

Proof According to the result in [5], the cost function c_1 is much better than continuous (in fact semiconcave) at (x, y) if the points x and y are not connected by abnormal minimizers (see [1] or below for the definitions of normal and abnormal minimizers). Therefore, let us apply Pontryagin maximum principle and find candidates for which the cost function c_1 is not continuous.

Let H_u^v be the Hamiltonian function defined by

$$H_u^v(x, p) = p(F(x, u)) + vL(x, u).$$

By applying Pontryagin maximum principle (see, for instance, [1]), any minimizer $(x(\cdot), u(\cdot))$ of the minimization problem in (1.3) satisfies

$$\dot{x}_i = \frac{\partial H_u^v}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H_u^v}{\partial x_i}, \quad \frac{\partial H_u^v}{\partial u_i} = 0 \tag{2.5}$$

for some curve $p(\cdot)$ and some constant v such that $(v, p(t)) \neq 0$. Moreover, v can be chosen to be either 0 or -1 . A minimizer $(x(\cdot), u(\cdot))$ is abnormal if the corresponding v in the Pontryagin maximum principle is 0. It is normal if $v = -1$. Note that a minimizer can both be normal and abnormal.

In the present case, the Hamiltonian H_u^v is given by

$$H_u^v(x, p) = p_1 u_1 + p_2 x_1^2 + p_2 u_2 x_1^k + \frac{v}{2} (u_1^2 + u_2^2).$$

For the abnormal case $v = 0$, (2.5) becomes

$$\dot{x}_1 = u_1, \quad \dot{x}_2 = x_1^2 + u_2 x_1^k, \quad \dot{p}_2 = 0, \quad p_1 = 0, \quad p_2 x_1^k = 0.$$

By Pontryagin maximum principle, p_1 and p_2 cannot be equal to zero simultaneously since $v = 0$. It follows that $x_1 \equiv 0$ and $x_2 \equiv x_2(0)$. The corresponding controls to all these paths are all given by the zero control $u \equiv 0$. It follows that $c_1((0, w), (0, w)) = 0$ for any w and these are candidates for discontinuities of the cost c_1 .

Next, we show that $c_1((0, w), (0, z)) \geq K$ for some constant $K > 0$ and for all $z < w$. For this, we consider the case $v = -1$. In this case, the Hamiltonian is given by

$$H_u^{-1}(x, p) = p_1 u_1 + p_2 x_1^2 + p_2 u_2 x_1^k - \frac{1}{2} (u_1^2 + u_2^2).$$

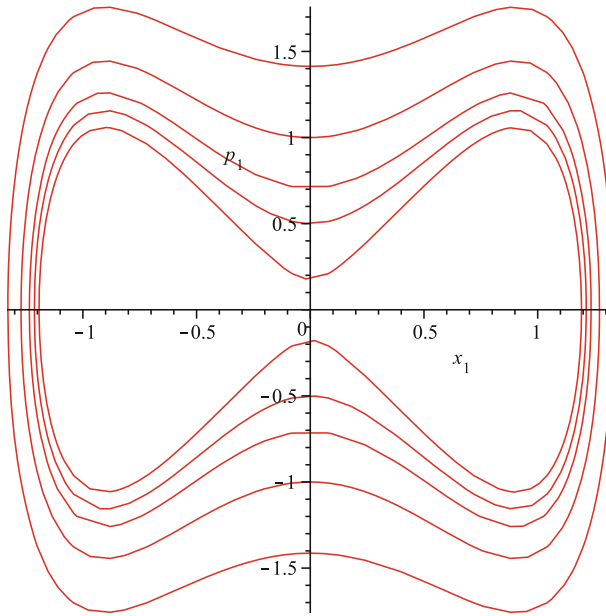


Fig. 1 Phase portrait of the example (level sets of the Hamiltonian H)

It follows from (2.5) that we have

$$\begin{aligned}
 H &:= H_u^{-1}(x, p) = \frac{1}{2}p_1^2 + \frac{1}{2}x_1^{2k}p_2^2 + x_1^2p_2, \\
 \dot{x}_1 &= p_1, \quad \dot{p}_2 = 0, \quad u_1 = p_1, \quad u_2 = x_1^k p_2.
 \end{aligned}
 \tag{2.6}$$

If we assume that $x(0) = (0, w)$ and $x(1) = (0, z)$ with $z < w$, then it follows from (2.6) that the cost $c_1((0, w), (0, z))$ for going from $(0, w)$ to $(0, z)$ is estimated by

$$c_1((0, w), (0, z)) = \frac{1}{2} \int_0^1 p_1^2 + x_1^{2k} p_2^2 dt \geq \frac{1}{2} \int_0^1 p_1^2 dt.
 \tag{2.7}$$

Since p_2 is a constant of motion, we can fix p_2 and look at the phase portrait of the system (see Fig. 1)

$$\dot{x}_1 = p_1, \quad \dot{p}_1 = -kx_1^{2k-1} p_2^2 - 2x_1 p_2.$$

The cost $c((0, w), (0, z))$ in (2.7) can be estimated from below by the area enclosed by the level set $H = 0$. More precisely,

$$c((0, w), (0, z)) \geq \int_0^\kappa p_1(x_1, p_2) dx_1,
 \tag{2.8}$$

where $p_1(x, p_2)$ is defined implicitly by $\frac{1}{2}p_1^2 + \frac{1}{2}x_1^{2k}p_2^2 + x_1^2p_2 = 0$ and $\kappa = \left(-\frac{2}{p_2}\right)^{\frac{1}{2k-2}}$ is the positive zero of the function $p_1(x_1, p_2)$. Note that $p_2 < 0$. Indeed, $H(x(t), p(t))$ is constant, we have $H(x(t), p(t)) = H_u^{-1}(x(0), p(0)) \geq 0$. It follows that $p_2(t) = p_2(0) \leq 0$.

If we do a change of variable $x_1 = \kappa z$, then we have

$$\begin{aligned} \int_0^\kappa p_1(x_1, p_2) dx_1 &\geq \int_0^\kappa \left(-2x_1^2 p_2 - x_1^{2k} p_2^2\right)^{\frac{1}{2}} dx_1 \\ &= 2^{\frac{k+1}{2k-2}} (-p_2)^{\frac{k-3}{2k-2}} \int_0^1 (z^2 - z^{2k})^{\frac{1}{2}} dz. \end{aligned} \tag{2.9}$$

On the other hand, by Fig. 1 and (2.6), we have

$$\frac{1}{2} \int_0^1 p_1^2 dt \geq \frac{1}{2} \int_0^1 |p_1| dt \geq \frac{1}{2} \int_{\{t|\dot{x}_1(t) \geq 0\}} \dot{x}_1 dt \geq \frac{1}{2} \left(-\frac{2}{p_2}\right)^{\frac{1}{2k-2}}. \tag{2.10}$$

If we combine (2.7), (2.8), (2.9), and (2.10), then we get

$$c((0, w), (0, z)) \geq C \max \left\{ \left(-\frac{1}{p_2}\right)^{\frac{1}{2k-2}}, (-p_2)^{\frac{k-3}{2k-2}} \right\}$$

for some constant $C > 0$.

It follows that the cost $c((0, w), (0, z))$ is bounded below by a positive constant independent of p_2 if $k \geq 3$ and this finishes the proof of the result. \square

3 Continuity of optimal control costs

In this section, we will state and prove the general continuity result (Theorem 3.2) mentioned in the Sect. 1. To do this, let us introduce some notations. If X_t is a, possibly time-dependent, vector field, then the corresponding flow φ_t defined by $\varphi_0(x) = x$ and $\frac{d}{dt}\varphi_t(x) = X_t(\varphi_t(x))$ is denoted by

$$\varphi_t = \overrightarrow{\text{exp}} \int_0^t X_s ds.$$

We define the endpoint map $\text{End}_{x_0}^T : L^P([0, T], U) \rightarrow M$ by

$$\text{End}_{x_0}^T(u(\cdot)) = \overrightarrow{\text{exp}} \int_0^T F_{u(s)} ds(x_0),$$

where F_u is the vector field defined by $F_u(x) = F(x, u) = X_0(x) + \sum_{i=1}^n u_i X_i(x)$.

Let us first fixed a control $u(\cdot)$. The first goal is to show that the control system is locally controllable. It means that we can reach any point near the point $\text{End}_{x_0}^T(u(\cdot))$ by adding a small control $v(\cdot)$ to the fixed one $u(\cdot)$. The first idea is to replace the control system (1.2) with drift X_0 by one without drift. However, the control vector fields X_1, \dots, X_n will become time dependent in the new control system. This is accomplished in Lemma 3.1. Recall that if $P : M \rightarrow M$ is a diffeomorphism and X is a vector field on M , then the pull back vector field P^*X is the vector field defined by $P^*X = dP^{-1}(X \circ P)$.

Lemma 3.1 *Let g_i^t be the time-dependent vector field defined by*

$$g_i^t := \left(\overrightarrow{\text{exp}} \int_0^t F_{u(s)} ds\right)^* X_i.$$

Then

$$End_{x_0}^T(u(\cdot) + v(\cdot)) = \overrightarrow{\text{exp}} \int_0^T F_{u(t)} dt \circ \overrightarrow{\text{exp}} \int_0^T \sum_{i=1}^n v_i(t) g_i^t dt(x_0).$$

Proof Let Q_t and R_t be the flows $\overrightarrow{\text{exp}} \int_0^t F_{u(s)+v(s)} ds$ and $\overrightarrow{\text{exp}} \int_0^t F_{u(s)} ds$, respectively. Let P_t be the flow defined by $Q_t = R_t \circ P_t$. If we differentiate the above equation, then we get

$$F_{u(t)+v(t)} \circ Q_t = \dot{Q}_t = \dot{R}_t \circ P_t + dR_t(\dot{P}_t) = F_{u(t)} \circ Q_t + dR_t(\dot{P}_t).$$

After simplifying the above equation, we get $\dot{P}_t = dR_t^{-1}(F_{v(t)} \circ Q_t) = (R_t)^* F_{v(t)} \circ P_t$ and this completes the proof. \square

Recall that we want to show local controllability by varying $v(\cdot)$. In Lemma 3.1, we have decompose the endpoint map $End_{x_0}(u(\cdot) + v(\cdot))$ into two parts. The first part $\overrightarrow{\text{exp}} \int_0^T F_{u(t)} dt$ is independent of the varying control $v(\cdot)$ and it is a diffeomorphism. Therefore, it is enough to show local controllability for the second term $\overrightarrow{\text{exp}} \int_0^T \sum_{i=1}^n v_i(t) g_i^t dt(x_0)$ which is the endpoint map to a new control system

$$\dot{x} = \sum_{i=1}^n v_i(t) g_i^t. \tag{3.11}$$

Note that this is a system with no drift but with time dependent control vector fields g_i^t as mentioned earlier.

Before proceeding to the proof of local controllability of the system (3.11), let us state the main result of this section which includes Theorem 1.2 as a corollary.

Theorem 3.2 *Assume that the Lagrangian L and the family of vector fields*

$$\{g_1^t, \dots, g_n^t \mid t \in [0, T]\}$$

satisfy the following conditions:

- (1) $C_1|u|^q + K_1 \leq L(x, u) \leq C_2|u|^p + K_2$,
- (2) $|\frac{\partial L(x, u)}{\partial x}| \leq C_3|u|^2$,
- (3) *the Hessian of L in the u variable is positive definite, and*
- (4) $\{g_1^t, \dots, g_n^t \mid t \in [0, T]\}$ *is k -generating, $\forall u(\cdot)$,*

for some constants $C_1, C_2, C_3, K_1, K_2 > 0$ and some constant $q > 1$. Suppose further that one of the followings is satisfied:

- (1) $k = 3$ and $p \leq 2$, or
- (2) $k > 3, p < \frac{k-2}{k-3}$.

Then the cost function $(t, x, y) \mapsto c_t(x, y)$ defined in (1.3) is continuous.

Going back to the local controllability issue of the system (3.11), let us denote the endpoint map to the new system by $\Phi^T : L^p([0, T], U) \times M \rightarrow M$. More precisely,

$$\Phi^T(v(\cdot), x) := \overrightarrow{\text{exp}} \int_0^T \sum_{i=1}^n v_i(t) g_i^t dt(x).$$

If the control vector fields g_i^t in the above new system is time independent, then local controllability follows from the Chow–Rashevskii theorem (see for instance [13]). More

precisely, we will need the following lemma for which the proof will be given for completeness. Recall that if X, Y are two vector fields, then the vector field $ad_X Y$ is defined by $ad_X Y = [X, Y]$.

Lemma 3.3 *Let g_1, \dots, g_n which are time-independent family of vector fields. Then there exists piecewise constant control $w(\cdot)$ for which $w(t)$ has only one nonzero component for each t and such that*

$$\begin{aligned}
 f\left(\overrightarrow{\text{exp}} \int_0^T \epsilon \left(\sum_{i=1}^n w_i(t) g_i\right) dt(x_0)\right) \\
 = f(x_0) + \epsilon^k (ad_{g_1} \dots ad_{g_{k-1}} g_k) f(x_0) + o(\epsilon^k)
 \end{aligned}
 \tag{3.12}$$

as $\epsilon \rightarrow 0$ for every smooth function f .

Proof Let P_t^ϵ and Q_t^ϵ be the flows corresponding to the control system (1.2) with controls ϵw^P and ϵw^Q , respectively. More precisely,

$$P_t^\epsilon(x_0) = \overrightarrow{\text{exp}} \int_0^t \epsilon \left(\sum_{i=1}^n w_i^P(s) g_i\right) ds, \quad Q_t^\epsilon(x_0) = \overrightarrow{\text{exp}} \int_0^t \epsilon \left(\sum_{i=1}^n w_i^Q(s) g_i\right) ds.$$

Moreover, assume that there are vector fields X and Y such that the flows P_t^ϵ and Q_t^ϵ satisfy

$$f(P_t^\epsilon(x_0)) = f(x_0) + \epsilon Xf(x_0) + o(\epsilon), \quad f(Q_t^\epsilon(x_0)) = f(x_0) + \epsilon^k Yf(x_0) + o(\epsilon^k)$$

for all smooth functions f .

Next, we define a control \bar{w} which is the concatenation of the controls $w^P, w^Q, -w^P$, and $-w^Q$.

$$\bar{w}(t) = \begin{cases} -w^P(t) & \text{if } 0 \leq t \leq T \\ -w^Q(t - T) & \text{if } T < t \leq 2T \\ w^P(t - 2T) & \text{if } 2T < t \leq 3T \\ w^Q(t - 3T) & \text{if } 3T < t \leq 4T. \end{cases}$$

It follows that

$$f\left(\overrightarrow{\text{exp}} \int_0^{4T} \epsilon \left(\sum_{i=1}^n \bar{w}_i(s) g_i\right) ds(x_0)\right) = f(Q_T^\epsilon \circ P_T^\epsilon \circ (Q_T^\epsilon)^{-1} \circ (P_T^\epsilon)^{-1}(x_0)).$$

Let $h(\epsilon_1, \epsilon_2) = f(Q_T^{\epsilon_2} \circ P_T^{\epsilon_1} \circ (Q_T^{\epsilon_2})^{-1} \circ (P_T^{\epsilon_1})^{-1}(x_0))$ and we want to consider the expansion of the function $h(\epsilon, \epsilon)$ in the parameter ϵ . Note that $P_T^0 = Q_T^0$ is the identity transformation. It follows that the zeroth order term of the expansion of $h(\epsilon, \epsilon)$ in ϵ is $f(x_0)$. In fact, the following is true.

$$h(\epsilon_1, 0) = h(0, \epsilon_2) = f(x_0).
 \tag{3.13}$$

By definition of the flow Q_t^ϵ , we have $\partial_\epsilon^i f(Q_t^\epsilon) \Big|_{\epsilon=0} = 0$ for all $i = 1, \dots, k - 1$. It follows that $\partial_{\epsilon_2}^i h \Big|_{\epsilon_2=0} = 0$ for each such i . Therefore, except the zeroth order term, any term of order less than k in the expansion of h vanishes. However, by (3.13), the k -th order vanishes as well. Therefore, we consider the $(k + 1)$ -th order term. Moreover, by the same argument, the

only nontrivial $(k + 1)$ -th order term is given by $\partial_{\epsilon_1} \partial_{\epsilon_2}^k h \Big|_{\epsilon_1=\epsilon_2=0}$. A computation shows the following

$$\partial_{\epsilon}^{k+1} h(\epsilon, \epsilon) \Big|_{\epsilon=0} = (k + 1) \partial_{\epsilon_1} \partial_{\epsilon_2}^k h(\epsilon_1, \epsilon_2) \Big|_{\epsilon_1=\epsilon_2=0} = (k + 1)[Y, X]f(x_0).$$

In conclusion, we have shown that

$$f\left(\overrightarrow{\text{exp}} \int_0^{4T} \epsilon \left(\sum_{i=1}^n \bar{w}_i(s) g_i\right) ds(x_0)\right) = f(x_0) + \frac{\epsilon^{k+1}}{k!} [Y, X]f(x_0) + o(\epsilon^{k+1}).$$

By rescaling time and multiplying the control \bar{w} by a constant, we have a control w which satisfies

$$f\left(\overrightarrow{\text{exp}} \int_0^T \epsilon \left(\sum_{i=1}^n w_i(s) g_i\right) ds(x_0)\right) = f(x_0) + \epsilon^{k+1} [Y, X]f(x_0) + o(\epsilon^{k+1}). \tag{3.14}$$

Note that if the controls w^P and w^Q are piecewise constant and have only one nonzero component for each time t , then so is w by construction.

If we let the control w^P and w^Q be the constant controls defined by $w_i^P(t) = \delta_{i,i_1}$ and $w_i^Q(t) = \delta_{i,i_2}$ for each t , then (3.14) shows that

$$f\left(\overrightarrow{\text{exp}} \int_0^T \epsilon \left(\sum_{i=1}^n w_i(s) g_i\right) ds(x_0)\right) = f(x_0) + \epsilon^2 [g_{i_2}, g_{i_1}]f(x_0) + o(\epsilon^2).$$

This proves the lemma for the case $k = 2$. The rest follows from induction using (3.14). \square

The second idea is to take a control given by Lemma 3.3, rescale it so that it is concentrated on a smaller and smaller time interval, and put the rescaled controls to the place where the vector fields g_1^t, \dots, g_n^t are bracket generating. This way we obtain local controllability as in Chow–Rashevskii theorem. Here we need the conditions on the numbers k and p to make sure that the rescaled controls stay small. This second idea will be achieved in Proposition 3.4 below. To do this, let us consider the curves $t \mapsto g_i^t(x_0)$ contained in the tangent space $T_{x_0}M$. Let \mathcal{I} be an interval in $[0, T]$ with the property that any subinterval \mathcal{I}' contained in \mathcal{I} satisfies

$$\text{span}\{g_1^t(x_0), \dots, g_n^t(x_0) | t \in \mathcal{I}'\} = \text{span}\{g_1^t(x_0), \dots, g_n^t(x_0) | t \in \mathcal{I}\}.$$

Proposition 3.4 *Let τ be a Lebesgue point of the control $u(\cdot)$ contained in the interval \mathcal{I} and assume that either*

- (1) $k = 3$ and $p \leq 2$, or
- (2) $k > 3$, $p < \frac{k-2}{k-3}$.

Then there exists $\alpha, \beta > 0$ and a family of controls $v^\epsilon(\cdot)$ which converges to 0 in L^p and such that

$$f(\Phi^T(v^\epsilon(\cdot)), x_0) = f(x_0) + \epsilon^{k(\beta-\alpha)} \int_0^T \text{ad}_{g_{i_1}^\tau} \dots \text{ad}_{g_{i_{k-1}}^\tau} (g_{i_k}^\tau) f(x_0) ds + o(\epsilon^{k(\beta-\alpha)}),$$

as $\epsilon \rightarrow 0$, for any smooth function f .

Proof By Lemma 3.3, there is a piecewise constant control $w(\cdot)$ for which $w(t)$ has only one nonzero component for each t and such that

$$\begin{aligned}
 & f\left(\overrightarrow{\text{exp}} \int_0^T \epsilon \left(\sum_{i=1}^n w_i(t) g_i^\tau \right) dt(x_0)\right) \\
 &= f(x_0) + \epsilon^k (ad_{g_{i_1}^\tau} \dots ad_{g_{i_{k-1}}^\tau} g_{i_k}^\tau) f(x_0) + o(\epsilon^k)
 \end{aligned}
 \tag{3.15}$$

as $\epsilon \rightarrow 0$. Note that τ is fixed and g_i^τ is a time independent vector field.

Let $0 = t_0 \leq t_1 \leq \dots \leq t_l = T$ be a partition such that the restriction $w|_{[t_{i-1}, t_i]}$ of the control $w(\cdot)$ to the subinterval $[t_{i-1}, t_i]$ is constant and there is only one nonzero component. We suppose that the k_i -th component of $w|_{[t_{i-1}, t_i]}$ is nonzero and this nonzero component is equal to c_i .

We need to create more freedom in our controls for later use (Lemma 3.5 to be precise). Let $v(\cdot)$ be a control of the form $v(\cdot) = w(\cdot) + \alpha(\cdot)$ such that $\alpha_j|_{[t_i, t_{i+1}]} \equiv 0$ if $j \neq k_i$ and $\int_{t_i}^{t_{i+1}} \alpha_{k_i}(s) ds = 0$. It follows from (3.15) and $\int_{t_i}^{t_{i+1}} \alpha_{k_i}(s) ds = 0$ that

$$\begin{aligned}
 & f\left(\overrightarrow{\text{exp}} \int_0^T \epsilon \left(\sum_{i=1}^n v_i(t) g_i^\tau \right) dt(x_0)\right) \\
 &= f(x_0) + \epsilon^k (ad_{g_{i_1}^\tau} \dots ad_{g_{i_{k-1}}^\tau} g_{i_k}^\tau) f(x_0) + o(\epsilon^k).
 \end{aligned}
 \tag{3.16}$$

Next, we rescale the control $v(\cdot)$ as mentioned earlier. Let $G_{s,v} := \sum_{i=1}^n v_i g_i^s$ and let

$$v^\epsilon(t) = \begin{cases} \epsilon^{-\alpha} v((t - \tau)/\epsilon^\beta) & \text{if } t \in (\tau, \tau + \epsilon^\beta) \\ 0 & \text{otherwise.} \end{cases}$$

Then we have

$$\begin{aligned}
 & f\left(\overrightarrow{\text{exp}} \int_0^T G_{s,v^\epsilon(s)} ds(x_0)\right) \\
 &= f\left(\overrightarrow{\text{exp}} \int_\tau^{\tau+\epsilon^\beta} G_{s,v^\epsilon(s)} ds(x_0)\right) \\
 &= f\left(\overrightarrow{\text{exp}} \int_0^T \epsilon^{\beta-\alpha} \sum_{i=1}^n v_i(s) g_i^{\epsilon^\beta s + \tau} ds(x_0)\right) \\
 &= f\left(\overrightarrow{\text{exp}} \int_0^T \epsilon^{\beta-\alpha} G_{\epsilon^\beta s + \tau, v(s)} ds(x_0)\right).
 \end{aligned}$$

By using the asymptotic expansion in [1, Sect. 2.4.4], the above equation becomes

$$\begin{aligned}
 & f\left(\overrightarrow{\text{exp}} \int_0^T G_{s,v^\epsilon(s)} ds(x_0)\right) \\
 &= f(x_0) + \sum_{i=1}^k \int_{0 \leq s_1 \leq \dots \leq s_i \leq T} \epsilon^{i(\beta-\alpha)} G_{\epsilon^\beta s_1 + \tau, v(s_1)} \dots G_{\epsilon^\beta s_i + \tau, v(s_i)} f(x_0) ds_1 \dots ds_i \\
 &+ o(\epsilon^{k(\beta-\alpha)}).
 \end{aligned}
 \tag{3.17}$$

as $\epsilon \rightarrow 0$.

Let I_i be the term

$$I_i(v(\cdot)) := \int_{0 \leq s_1 \leq \dots \leq s_i \leq T} G_{\epsilon^{\beta s_1 + \tau, v(s_1)}} \dots G_{\epsilon^{\beta s_i + \tau, v(s_i)}} f(x_0) ds_1 \dots ds_i$$

in the expansion (3.17).

Let us first deal with the term $I_1(v(\cdot)) = \int_0^T G_{\epsilon^{\beta s + \tau, v(s)}} f(x_0) ds$. For this, let $g_i^{\tau + \epsilon s} = g_i^\tau + Z_i^{\epsilon, s}$. Let us recall that $v(\cdot) = w(\cdot) + \alpha(\cdot)$ and $\alpha_j \equiv 0$ if $j \neq k_i$.

$$\begin{aligned} I_1(v(\cdot)) &= \int_0^T \sum_{i=1}^n v_i(s) g_i^{\tau + \epsilon \beta s} f(x_0) ds \\ &= \int_0^T \sum_{i=1}^n v_i(s) g_i^\tau f(x_0) ds + \int_0^T \sum_{i=1}^n (w_i(s) + \alpha_i(s)) Z_i^{\epsilon, s} f(x_0) ds \\ &= \int_0^T \sum_{i=1}^n v_i(s) g_i^\tau f(x_0) ds + \sum_{i=1}^n \int_{t_{i-1}}^{t_i} (c_i + \alpha_{k_i}(s)) Z_{k_i}^{\epsilon, s} f(x_0) ds. \end{aligned} \tag{3.18}$$

The next lemma says that we can choose α to get rid of the last term of the above equation.

Lemma 3.5 *There exists $\epsilon_0 > 0$ and $\alpha^\epsilon(\cdot)$ in L^2 such that $\int_{t_{i-1}}^{t_i} \alpha_{k_i}^\epsilon(s) ds = 0$ and*

$$I_1(w(\cdot) + \alpha^\epsilon(\cdot)) = \int_0^T \sum_{i=1}^n v_i(s) g_i^\tau f(x_0) ds \tag{3.19}$$

for all $0 < \epsilon < \epsilon_0$. Moreover, α^ϵ goes to 0 in L^2 as ϵ goes to 0.

Proof of Lemma 3.5 Recall that we need $\alpha^\epsilon(\cdot)$ to satisfy the conditions

$$\int_{t_{i-1}}^{t_i} \alpha_{k_i}^\epsilon(s) ds = 0, \quad \int_{t_{i-1}}^{t_i} (c_i + \alpha_{k_i}^\epsilon(s)) Z_{k_i}^{\epsilon, s} f(x_0) ds = 0. \tag{3.20}$$

for all smooth functions f and for all i . Consider local coordinates around the point x_0 and suppose that $Z_{k_i}^{\epsilon, s} = (Z_{k_i, 1}^{\epsilon, s}, \dots, Z_{k_i, m}^{\epsilon, s})$ in this local coordinates. Then the conditions in (3.20) is the same as that $\alpha_{k_i}^\epsilon(\cdot)$ orthogonal to the constant functions and $c_i + \alpha_{k_i}^\epsilon(\cdot)$ is orthogonal to $Z_{k_i}^{\epsilon, \cdot}$ in $L^2([t_{i-1}, t_i])$ for each i . Let V_i be the finite dimensional subspace of $L^2([t_{i-1}, t_i])$ defined by

$$V_i^\epsilon := \text{span}\{Z_{k_i, 1}^{\epsilon, \cdot}, \dots, Z_{k_i, m}^{\epsilon, \cdot}\}.$$

A linear algebra argument shows that $\alpha^\epsilon(\cdot)$ which satisfy the conditions (3.20) exist if V_i^ϵ does not contain any nonzero constant function. Therefore, we assume that

$$\sum_{j=1}^m a_j Z_{k_i, j}^{\epsilon, s} = c$$

for some constants a_1, \dots, a_m, c , for all s in $[t_{i-1}, t_i]$, and for some i . We are going to show that c must be zero and this finishes the proof of the lemma.

The above equation means that $(Z_{k_i, 1}^{\epsilon, s}, \dots, Z_{k_i, m}^{\epsilon, s})$ is contained in the affine space $\{z \in \mathbb{R}^m \mid \sum_{i=1}^m a_i z_i = c\}$ for almost all s in the interval $[t_{i-1}, t_i]$. Therefore, $(\frac{d}{ds} Z_{k_i, 1}^{\epsilon, s}, \dots, \frac{d}{ds} Z_{k_i, m}^{\epsilon, s})$ is contained in the subspace $\{z \in \mathbb{R}^m \mid \sum_{i=1}^m a_i z_i = 0\}$ for each s in the interval $[t_{i-1}, t_i]$. Let us choose ϵ_0 such that $\tau + t\epsilon$ is contained in the interval \mathcal{I} for each

t in $[0, T]$ and for all $\epsilon < \epsilon_0$. Then it follows from the definition of the interval \mathcal{I} that $(\frac{d}{ds}Z_1^{\epsilon,s}(x_0), \dots, \frac{d}{ds}Z_m^{\epsilon,s}(x_0))$ is contained in $\{z \in \mathbb{R}^m \mid \sum_{i=1}^m a_i z_i = 0\}$ for almost all s in $[0, T]$. Therefore, $(Z_{k_i,1}^{\epsilon,s}(x_0), \dots, Z_{k_i,m}^{\epsilon,s}(x_0))$ is contained in the affine space $\{z \in \mathbb{R}^m \mid \sum_{i=1}^m a_i z_i = c\}$ for all s in $[0, T]$. However, $(Z_{k_i,1}^{\epsilon,0}, \dots, Z_{k_i,m}^{\epsilon,0}) = 0$, so $c = 0$ and this finishes the proof of the lemma. \square

For the rest of the proof, we write $v(\cdot) = w(\cdot) + \alpha^\epsilon(\cdot)$ and suppress the ϵ -dependence on v to avoid complicated notation.

Lemma 3.6

$$I_k(v(\cdot)) = \int_{0 \leq s_1 \leq \dots \leq s_k \leq T} \sum_{i_1, \dots, i_k=1}^n v_{i_1}(s_1) \dots v_{i_k}(s_k) g_{i_1}^\tau \dots g_{i_k}^\tau f(x_0) ds_1 \dots ds_k + o(1)$$

as $\epsilon \rightarrow 0$.

Proof of Lemma 3.6 This follows immediately from the definition of $G_{\tau,v}$. Indeed,

$$\begin{aligned} I_k(v(\cdot)) &:= \int_{0 \leq s_1 \leq \dots \leq s_k \leq T} G_{\epsilon^\beta s_1 + \tau, v(s_1)} \dots G_{\epsilon^\beta s_k + \tau, v(s_k)} f(x_0) ds_1 \dots ds_k \\ &= \int_{0 \leq s_1 \leq \dots \leq s_k \leq T} \sum_{i_1, \dots, i_k=1}^n v_{i_1}(s_1) \dots v_{i_k}(s_k) g_{i_1}^{\tau + \epsilon^\beta s_1} \dots g_{i_k}^{\tau + \epsilon^\beta s_k} f(x_0) ds_1 \dots ds_k \\ &= \int_{0 \leq s_1 \leq \dots \leq s_k \leq T} \sum_{i_1, \dots, i_k=1}^n v_{i_1}(s_1) \dots v_{i_k}(s_k) g_{i_1}^\tau \dots g_{i_k}^\tau f(x_0) ds_1 \dots ds_k + o(1). \end{aligned} \tag{3.21}$$

\square

If we combine Lemma 3.5 and Lemma 3.6 with (3.17) and assume that $3\beta - 2\alpha > k(\beta - \alpha) > 0$, then we have

$$\begin{aligned} &f\left(\overrightarrow{\text{exp}} \int_0^T G_{s, v^\epsilon(s)} ds(x_0)\right) \\ &= f(x_0) + \sum_{i=1}^k \epsilon^{i(\beta-\alpha)} I_i(w(\cdot) + \alpha^\epsilon(\cdot)) + O\left(\epsilon^{k(\beta-\alpha)}\right) \\ &= f(x_0) + \sum_{j=1}^k \epsilon^{j(\beta-\alpha)} \int_{0 \leq s_1 \leq \dots \leq s_j \leq T} \sum_{i_1, \dots, i_j=1}^n v_{i_1}(s_1) \dots v_{i_j}(s_j) g_{i_1}^\tau \dots g_{i_j}^\tau f(x_0) ds_1 \dots ds_j \\ &\quad + o\left(\epsilon^{k(\beta-2\alpha)}\right) \end{aligned}$$

as $\epsilon \rightarrow 0$.

By (3.16), the above becomes

$$\begin{aligned} & f\left(\overrightarrow{\text{exp}} \int_0^T G_{s, v^\epsilon(s)} ds(x_0)\right) \\ &= f\left(\overrightarrow{\text{exp}} \int_0^T \epsilon^{\beta-\alpha} \sum_{i=1}^n v_i(t) g_i^\tau dt\right) + o\left(\epsilon^{k(\beta-\alpha)}\right) \\ &= f(x_0) + \epsilon^{k(\beta-\alpha)} \left(ad_{g_{i_1}}^\tau \dots ad_{g_{i_{k-1}}} g_{i_k}\right) f(x_0) + o\left(\epsilon^{k(\beta-\alpha)}\right). \end{aligned}$$

Finally, we need $v^\epsilon(\cdot)$ converges to 0 in L^p . Indeed, by the definition of $v^\epsilon(\cdot)$, we have

$$\begin{aligned} \int_0^T |v^\epsilon(t)|^p dt &= \int_\tau^{\tau+\epsilon^\beta} \left| \epsilon^{-\alpha} v\left(\frac{t-\tau}{\epsilon^\beta}\right) \right|^p dt \\ &= \int_0^1 |v(s)|^p \epsilon^{\beta-\alpha p} ds \\ &= \int_0^1 |w(s) + \alpha^\epsilon(s)|^p \epsilon^{\beta-\alpha p} ds. \end{aligned}$$

Since $w(\cdot)$ is in L^∞ and $\alpha^\epsilon(\cdot)$ is in L^2 , $v^\epsilon(\cdot)$ converges to 0 in L^p if $\beta - \alpha p > 0$ and $p \leq 2$.

In conclusion, if we can choose α and β such that the following three conditions are satisfied, then the conclusion of the theorem holds.

$$3\beta - 2\alpha > k(\beta - \alpha) > 0, \quad \beta - \alpha p > 0, \quad p \leq 2.$$

It is not hard to check that these inequalities are satisfied under the assumptions of the proposition. □

The local controllability of the control system follows using Proposition 3.4 and implicit function theorem as in the Chow–Rashevskii theorem. Finally, the continuity of the cost follows from the local controllability and standard arguments as in [5].

Proof of Theorem 3.2 Lower semi-continuity of the cost can be proved in the same way as in [5]. To prove upper semi-continuity, we let $(x_1, y_1, t_1), (x_2, y_2, t_2), \dots$ be a sequence of points which converges to (x, y, T) and $\lim_{i \rightarrow \infty} c_{t_i}(x_i, y_i) = r$. We want to show that $c_T(x, y) \geq r$.

Assume that this is not the case. Let $u(\cdot)$ and $x(\cdot)$ be a control and the trajectory associated to this control, respectively, such that $x(0) = x, x(T) = y$ and $\int_0^T L(x(s), u(s)) ds < r$. Recall that the family of vector fields $\{g_1^t, \dots, g_n^t | t \in [0, T]\}$ is k -generating. Therefore, we can find vector fields V_1, \dots, V_k from the set

$$\left\{ ad_{g_{i_1}}^{t_1} \dots ad_{g_{i_l}}^{t_l} \mid 1 \leq i_j \leq n, 0 \leq t \leq T, 0 \leq l \leq k \right\}.$$

which span the tangent space $T_x M$. We also assume that V_i is defined by the Lie brackets of κ_i vector fields of the form $g_j^{\tau_j}$. By perturbation, we can assume $\tau_i \neq \tau_j$ for $i \neq j$ and that each τ_i satisfies the condition in Theorem 3.4. Therefore, by Theorem 3.4, there is a family of control $w_{i,\epsilon}(\cdot)$ such that

$$f\left(\Phi^T(w_{i,\epsilon}(\cdot), x_0)\right) = f(x_0) + \epsilon^{\kappa_i(\beta-\alpha)} \int_0^T V_i f(x_0) ds + o\left(\epsilon^{\kappa_i(\beta-\alpha)}\right).$$

Note that from the proof of Proposition 3.4, we can assume that w_{i,ϵ_i} is supported in a small interval J_i around τ_i by taking $(\epsilon_1, \dots, \epsilon_n)$ small enough. Moreover, we can assume that the intervals J_i are disjoint. We define the map $\Psi : M \times \mathbb{R}^n \times (0, \infty) \rightarrow M$ by

$$\begin{aligned} \Psi(x, \epsilon_1, \dots, \epsilon_n, T) &= \Phi^T(w_{1,\epsilon_1/\kappa_1(\beta-\alpha)}) \circ \Phi^T(w_{2,\epsilon_2/\kappa_2(\beta-\alpha)}) \\ &\times \dots \circ \Phi^T(w_{n,\epsilon_n/\kappa_n(\beta-\alpha)})(x). \end{aligned}$$

Since $\frac{d}{d\epsilon_i} \Big|_{\epsilon_i=0} \Psi = V_i$, the map Ψ is of full rank at the point $(x, 0, \dots, 0, T)$. It follows from implicit function theorem that there exists a map $\psi : U_1 \rightarrow U_2$ from a neighborhood U_1 of (x, y, T) to a neighborhood U_2 of $(0, \dots, 0)$ such that $\Psi(z_1, \psi(z_1, z_2, t), t) = z_2$ for all pairs (z_1, z_2, t) in the set U_1 .

Let $(\epsilon_1^i, \dots, \epsilon_n^i) = \psi(x_i, y_i, t_i)$ and let $v^i(\cdot)$ be the control defined by $v^i(t) = w_{j,\epsilon_j^i}(t)$ and 0 otherwise. $v^i(\cdot)$ is well defined if $(\epsilon_1^i, \dots, \epsilon_n^i)$ is close enough to $(0, \dots, 0)$. Let $x^i(\cdot)$ be a curve in M which satisfies (1.2) with control $v^i(\cdot)$. We know that $v^i(\cdot)$ converges strongly in L^p to 0 and $x^i(\cdot)$ converges uniformly to $x(\cdot)$.

Assume, without loss of generality, that $u(t) = 0$ for all $t > T$. Then

$$\begin{aligned} &\left| \int_0^{t_i} L(x^i(s), u(s) + v^i(s)) ds - \int_0^T L(x(s), u(s)) ds \right| \\ &\leq \left| \int_0^{t_i} L(x^i(s), u(s) + v^i(s)) ds - \int_0^{t_i} L(x(s), u(s) + v^i(s)) ds \right| \\ &\quad + \left| \int_0^{\tau_i} L(x(s), u(s) + v^i(s)) ds - \int_0^{\tau_i} L(x(s), u(s)) ds \right| \\ &\quad + \left| \int_{t_i}^T \max\{L(x(s), u(s) + v^i(s)), L(x(s), u(s))\} ds \right|, \end{aligned} \tag{3.22}$$

where $\tau_i = \min\{T, t_i\}$.

Since $L(x, u) \leq C_2|u|^p + K_2$, t_i converges to T , and the sequence $u(\cdot) + v^i(\cdot)$ converges to $u(\cdot)$ in L^p , we have

$$\begin{aligned} &\left| \int_{t_i}^T \max\{L(x(s), u(s) + v^i(s)), L(x(s), u(s))\} ds \right| \\ &\leq \int_{t_i}^T \max\{C_2|u(s) + v^i(s)|^p - K_2, C_2|u(s)|^p - K_2\} ds \\ &\rightarrow 0 \end{aligned} \tag{3.23}$$

as $i \rightarrow \infty$.

Recall that $|\frac{\partial L}{\partial x}| \leq C_3|u|^2$, where the norm is taken with respect to certain Riemannian metric. Let d be the corresponding Riemannian distance function. Then we have

$$\begin{aligned} &\left| \int_0^{t_i} L(x^i(s), u(s) + v^i(s)) ds - \int_0^{t_i} L(x(s), u(s) + v^i(s)) ds \right| \\ &\leq \int_0^{t_i} \left| L(x^i(s), u(s) + v^i(s)) - L(x(s), u(s) + v^i(s)) \right| ds \\ &\leq \sup_s d(x^i(s), x(s)) \int_0^{t_i} C_3 |u(s) + v^i(s)|^2 ds \\ &\rightarrow 0 \quad \text{as } i \rightarrow \infty. \end{aligned} \tag{3.24}$$

By construction of the control $v^i(\cdot)$, we know that the indicator function $\mathbb{I}_{\{t|v_i(t)\neq 0\}}$ converges to zero almost everywhere. It follows that

$$\begin{aligned} & \left| \int_0^{\tau_i} L(x(s), u(s) + v^i(s)) ds - \int_0^{\tau_i} L(x(s), u(s)) ds \right| \\ & \leq \int_0^{\tau_i} \mathbb{I}_{\{t|v^i(t)\neq 0\}} \left(\left| L(x(s), u(s) + v^i(s)) \right| + |L(x(s), u(s))| \right) ds \\ & \rightarrow 0 \quad \text{as } i \rightarrow \infty. \end{aligned} \tag{3.25}$$

Therefore, if we combine (3.22), (3.23), (3.24), and (3.25), then we have

$$\lim_{i \rightarrow \infty} \int_0^{\tau_i} L(x^i(s), u(s) + v^i(s)) ds = \int_0^T L(x(s), u(s)) ds < r.$$

On the other hand,

$$\lim_{i \rightarrow \infty} \int_0^{\tau_i} L(x^i(s), u(s) + v^i(s)) ds \geq \lim_{i \rightarrow \infty} c_{\tau_i}(x_i, y_i) = r.$$

Therefore, this gives a contradiction and we finish the proof of upper semi-continuity of the function $(t, x, y) \mapsto c_t(x, y)$. \square

4 Optimal control and weak KAM theorem

In this section, we give a proof of Theorem 1.3 using some ideas from [2] and [3]. More precisely, we will prove the following.

Theorem 4.1 *Assume that the function $(t, x, y) \mapsto c_t(x, y)$ defined by (1.3) is continuous and the manifold M is compact, then there exists a unique constant h such that the Hamilton–Jacobi–Bellman equation (1.1) has a viscosity solution.*

We start the proof by introducing the Lax–Oleinik semigroup:

$$S_t f(y) = \inf_{x \in M} [c_t(x, y) + f(x)]. \tag{4.26}$$

Theorem 4.2 *Assume that the function $(t, x, y) \mapsto c_t(x, y)$ defined by (1.3) is continuous. Then, for each function f , the function $(t, x) \mapsto S_t f(x)$ is continuous on $(0, \infty) \times M$. Moreover, it is a viscosity solution to the Hamilton–Jacobi–Bellman equation $\partial_t f + H(x, \partial_x f) = 0$ on $(0, \infty) \times M$.*

Proof Continuity of the function $S_t f$ follows immediately from that of c_t and compactness of the manifold M . The fact that it is a viscosity solution follows as in [7]. \square

The following theorem is a continuous version of [3, Lemma 9] and the proof is similar.

Theorem 4.3 *Assume that the function $(t, x, y) \mapsto c_t(x, y)$ is continuous and the manifold M is compact. Then, for each $a > 0$, the family $\{c_t | t \geq a\}$ is equicontinuous. Moreover, there exists constants h and K such that*

$$|c_t(x, y) - ht| \leq K$$

for all $t \geq a$ and all x, y in M .

Proof The function $(t, x, y) \mapsto c_t(x, y)$ is uniformly continuous on $[a, b] \times M \times M$ for some constants $b > 2a > 0$. So, given $\epsilon > 0$, there is a $\delta > 0$ such that

$$|c_\tau(x_1, y_1) - c_\tau(x_2, y_2)| < \epsilon/2$$

whenever $d(x_1, x_2) < \delta, d(y_1, y_2) < \delta$ and $a < \tau < b$.

Assume that $t \geq a$. Since $b > 2a$, there is a partition $0 = t_0 \leq t_1 \leq t_2 \leq \dots \leq t_l = t$ such that $a < t_{i+1} - t_i < b$. Assume that $c_t(x_1, y_1) \geq c_t(x_2, y_2)$ and let $x_2 = z_0, z_1, \dots, z_l = y_2$ be points on the manifold M such that $c_t(x_2, y_2) = \sum_{i=1}^l c_{t_i}(z_{i-1}, z_i)$. Then we have

$$\begin{aligned} c_t(x_1, y_1) - c_t(x_2, y_2) &\leq c_{t_1-t_0}(x_1, z_1) - c_{t_1-t_0}(x_2, z_1) + c_{t_2-t_1}(z_{l-1}, y_1) - c_{t_2-t_1}(z_{l-1}, y_2) \\ &< \epsilon \end{aligned}$$

whenever $d(x_1, x_2) < \delta$ and $d(y_1, y_2) < \delta$. It follows that $\{c_t | t \geq a\}$ is an equicontinuous family.

Let $M_t = \sup_{x,y} c_t(x, y)$ and $m_t = \inf_{x,y} c_t(x, y)$, where the supremum and the infimum are taken over all pairs of points of the manifold M . Let t_1 and t_2 be two positive numbers and let z be a point on the manifold M such that $c_{t_1+t_2}(x, y) = c_{t_1}(x, z) + c_{t_2}(z, y)$. It follows from this $M_{t_1+t_2} \leq M_{t_1} + M_{t_2}$. Similarly, m_t satisfies $m_{t_1+t_2} \geq m_{t_1} + m_{t_2}$. It follows that the infimum of the function $\frac{M_t}{t}$ is finite. Indeed, if the infimum of $\frac{M_t}{t}$ is $-\infty$, then so is $\frac{m_t}{t}$. But note that $\frac{m_{kt_0}}{kt_0} \geq \frac{m_{t_0}}{t_0}$ for all positive integer k . This gives a contradiction. It follows that $M := \inf_t \frac{M_t}{t}$ is finite. Given $\epsilon > 0$, we find t_0 such that $\frac{M_{t_0}}{t_0} < M + \epsilon$. Every $t > t_0$ can be decompose into $t = kt_0 + s$, where $t_0 \leq s \leq 2t_0$. It follows that

$$M \leq \frac{M_t}{t} \leq \frac{kM_{t_0} + M_s}{t} = \frac{M_{t_0}}{t_0} \frac{kt_0}{kt_0 + s} + \frac{M_s}{t} < (M + \epsilon) \frac{kt_0}{kt_0 + s} + \frac{M_s}{t}.$$

By continuity of the cost c , we know that M_s is bounded. It follows that from this and the above inequality that

$$\lim_{t \rightarrow \infty} \frac{M_t}{t} = M.$$

Similarly, we also have

$$\lim_{t \rightarrow \infty} \frac{m_t}{t} = m.$$

Finally, it follows from equicontinuity of the family $\{c_t | t \geq a\}$ that $M_t - m_t \leq C$ for some constant C and for all $t \geq a$. Therefore, $h := M = m$. □

Lemma 4.4 *Assume that the function $(t, x, y) \mapsto c_t(x, y)$ is continuous and the manifold M is compact. Let f be a bounded function, then the family $S := \{S_t f - ht | t \geq a\}$ is uniformly bounded and equicontinuous.*

Proof According to Lemma 4.3, the family $\{c_t | t \geq a\}$ is equicontinuous. So, for each $\epsilon > 0$, there is a $\delta > 0$ such that for all $t \geq a$

$$|c_t(x_1, y_1) - c_t(x_2, y_2)| < \epsilon/2$$

whenever $d(x_1, x_2) < \delta$ and $d(y_1, y_2) < \delta$.

By definition of $S_t f$, we can find, for each $\epsilon > 0$, a point z_t such that

$$S_t f(x_2) > c_t(z_t, x_2) + f(z_t) - \epsilon/2.$$

It follows that

$$S_t f(x_1) - S_t f(x_2) < c_t(z_t, x_1) + f(z_t) - c_t(z_t, x_2) - f(z_t) + \epsilon/2 < \epsilon.$$

Since the above equation holds for all ϵ and all $t \geq a$, we conclude that the family \mathcal{S} is equicontinuous.

Fix a point x in M . For each $\epsilon > 0$, let z be a point in M such that

$$c_t(z, x) + f(z) \geq S_t f(x) > c_t(z, x) + f(z) - \epsilon.$$

Therefore, by Theorem 4.3, we have

$$K + \sup_{y \in M} \{f(y)\} \geq S_t f(x) - ht > -K + \inf_{y \in M} \{f(y)\} - \epsilon$$

for some constant $K > 0$. We conclude from this that \mathcal{S} is uniformly bounded. □

Define the function \bar{f} by

$$\bar{f}(x) = \inf_{t \geq a} [S_t f(x) - ht].$$

It follows from Lemma 4.4 that \bar{f} is bounded. The following theorem taken from [9] together with Theorem 4.2 and Lemma 4.4 finish the proof of the existence part of Theorem 4.1. We give a sketch of the proof here.

Theorem 4.5 *Assume that there exists a constant h such that the family $\mathcal{S} := \{S_t f - ht | t \geq a\}$ is uniformly bounded and equicontinuous, then $S_t \bar{f} - ht$ converges uniformly to a function \tilde{f} . Moreover, it satisfies*

$$S_t \tilde{f} - ht = \tilde{f}.$$

Proof By applying S_t to the definition of \bar{f} , it is not hard to see that $S_t \bar{f} - ht \geq \bar{f}$. Since S_t is order preserving, we can apply S_t again to this inequality to shows that $t \mapsto S_t \bar{f}(x) - ht$ is increasing for each x in M . It follows from this and Lemma 4.4 that $S_t \bar{f}(x) - ht$ converges uniformly to a continuous function \tilde{f} . We apply once again S_t to the definition of \bar{f} and use the continuity of the semigroup S_t , we get $S_t \tilde{f} - kt = \tilde{f}$. □

Finally, we finish the uniqueness of the constant h as a corollary of Theorem 4.3.

Corollary 4.6 *Assume that the function $(t, x, y) \mapsto c_t(x, y)$ is continuous and the manifold M is compact. Let h be as in Theorem 4.3 and let f be a function which satisfies $S_t f - kt = f$ for some number k , then $k = h$.*

Proof For each natural number n , let z_n be points in M which satisfies

$$c_n(z_n, x) + f(z_n) \geq f(x) + kn = S_n f(x) \geq c_n(z_n, x) + f(z_n) - \frac{1}{n}.$$

Note that the function f is continuous and $\lim_{n \rightarrow \infty} \frac{c_n(z_n, x)}{n} = h$. It follows that if we divide the above inequality by n and let n goes to infinity, we get $k = h$ as claimed. □

5 Optimal transportation and weak KAM theorem

Let μ and ν be two Borel probability measures. Consider the cost function defined in (1.3) and the following Monge–Kantorovich problem of optimal transportation:

$$C_T(\mu, \nu) = \inf_{\Pi} \int_{M \times M} c_T(x, y) d\Pi(x, y) \tag{5.27}$$

where the infimum is taken over all measures on $M \times M$ with marginals μ and ν . That is, if $\pi_1, \pi_2 : M \times M \rightarrow M$ are the projections onto the first and second entries, then $\pi_{1*}\Pi = \mu$ and $\pi_{2*}\Pi = \nu$.

The above problem (5.27) admits a dual version given by

$$\mathcal{I}_T(\mu, \nu) = \sup \int_M g(x) d\nu(x) - \int_M f(x) \mu(x), \tag{5.28}$$

where the supremum is taken over all pairs of functions (f, g) which satisfy $g(y) - f(x) \leq c_T(x, y)$.

The following theorem is the well known result in [11]. See also [14, 15].

Theorem 5.1 *Assume that the function c_T is continuous, then the infimum in (5.27) and the supremum in (5.28) is achieved. Moreover, for any optimal measure Π of (5.27) and any pair of functions (f, g) that maximizes (5.28), we have that Π is concentrated on the set $\{(x, y) \in M \times M | g(y) - f(x) = c_T(x, y)\}$ and $C_T(\mu, \nu) = \mathcal{I}_T(\mu, \nu)$.*

Note that if (f, g) maximizes (5.28), then so is $(f, S_T f)$. We define

$$\alpha_T := \inf_{\mu} \frac{1}{T} C_T(\mu, \mu), \tag{5.29}$$

where the infimum is taken over all Borel probability measures on M .

The following lemma can be proved in same way as in [2, Lemma 33].

Lemma 5.2 *There exists a measure μ which achieves the infimum in (5.29).*

The next theorem is a generalization of a result [2] which gives another characterization of the number h in Theorem 1.3.

Theorem 5.3 *Under the assumptions in Theorem 1.3, we have $\alpha_T = h$ for each $T > 0$.*

Following [3], we call measures Π on the space $M \times M$ generalized Mather measure if $\pi_{1*}\Pi = \pi_{2*}\Pi$ and

$$\frac{1}{T} \int_{M \times M} c_T(x, y) d\Pi(x, y) = h.$$

The following corollary describes the support of the generalized Mather measures.

Corollary 5.4 *Suppose that we make the same assumptions as in Theorem 1.3 and let g be a function which satisfies $S_t g = g + ht$. If Π is a generalized Mather measure which satisfies*

$$\frac{1}{T} \int_{M \times M} c_T(x, y) d\Pi(x, y) = h,$$

then the support of Π is contained in the set

$$\{(x, y) | c_T(x, y) = g(y) - g(x) + hT\}.$$

Proof Let $\mu = \pi_{1*}\Pi = \pi_{2*}\Pi$. Then

$$C_T(\mu, \mu) = hT = \int_M S_T g d\mu - \int_M g d\mu \leq \mathcal{I}_T(\mu, \mu) = C_T(\mu, \mu).$$

It follows that the support of Π is contained in

$$\{(x, y) | c_T(x, y) = S_T g(y) - g(x)\}.$$

□

Proof of Theorem 5.3 Let g be a function which satisfies $S_t g = g + ht$ and let μ be a minimizer corresponding to the minimization problem of α_T in (5.29). It follows from Theorem 5.1 that

$$T\alpha_T = C_T(\mu, \mu) = \mathcal{I}_T(\mu, \mu) \geq \int_M S_T g d\mu - \int_M g d\mu = hT \tag{5.30}$$

for all $T > 0$.

For the proof of the following lemma, we follow closely [3, Lemma 7].

Lemma 5.5 *Let ν_1 and ν_2 be two Borel probability measures and let $0 \leq s \leq T$, then there exists a Borel probability measure ν such that*

$$C_T(\nu_1, \nu_2) = C_s(\nu_1, \nu) + C_{T-s}(\nu, \nu_2).$$

Proof By Theorem 5.1, we can find measures Π_1, Π_2 on $M \times M$ such that $C_s(\nu_1, \nu) = \int_{M \times M} c_s(x, y) d\Pi_1(x, y)$ and $C_{T-s}(\nu, \nu_2) = \int_{M \times M} c_{T-s}(x, y) d\Pi_2(x, y)$. By disintegration of measures, there are measures μ_1^y and μ_2^x such that $d\Pi_1(x, y) = d\mu_1^y(x) d\nu(y)$ and $d\Pi_2(x, y) = d\mu_2^x(y) d\nu(x)$. Let μ be the measure defined by

$$\int_{M \times M} f(x, y) d\Pi(x, y) = \int_{M \times M \times M} f(x, y) d\mu_1^z(x) d\mu_2^z(y) d\nu(z).$$

It is not hard to check that the marginals of Π are ν_1 and ν_2 . Therefore, we get

$$\begin{aligned} C_T(\nu_1, \nu_2) &\leq \int_{M \times M} c_T(x, y) d\Pi(x, y) \\ &\leq \int_{M \times M \times M} c_s(x, z) + c_{T-s}(z, y) d\mu_1^z(x) d\mu_2^z(y) d\nu(z) \\ &= \int_{M \times M} c_s(x, z) d\Pi_1(x, z) + \int_{M \times M} c_{T-s}(z, y) d\Pi_2(z, y) \\ &= C_s(\nu_1, \nu) + C_{T-s}(\nu, \nu_2). \end{aligned}$$

Let \mathcal{P} be the set of pairs of Borel probability measures (ν_1, ν_2) which satisfies the conclusion of the lemma. It is not hard to see that (δ_x, δ_y) is contained in \mathcal{P} , where δ_x is the Dirac mass at x . Indeed, let $x(\cdot) : [0, T] \rightarrow M$ be an admissible path which satisfy $x(0) = x, x(T) = y$ and achieve the infimum in (1.3). Then,

$$\begin{aligned} C_s(\delta_x, \delta_{x(s)}) + C_{T-s}(\delta_{x(s)}, \delta_y) &= c_s(x, x(s)) + c_{T-s}(x(s), y) \\ &= c_T(x, y) = C_T(\delta_x, \delta_y). \end{aligned} \tag{5.31}$$

To finish the proof, it remains to notice that the set \mathcal{P} is convex and weak-* closed. Therefore, the result follows from approximation by delta masses. □

Now let ν be a measure which satisfies $C_{NT}(\nu, \nu) = \alpha_{NT}$. It follows from Lemma 5.5 that there exists Borel probability measures $\nu = \mu_0, \mu_1, \dots, \mu_N = \nu$ such that

$$C_{NT}(\nu, \nu) = \sum_{i=1}^N C_T(\mu_{i-1}, \mu_i).$$

Since \mathcal{I}_T is convex and so is C_T . Therefore,

$$\frac{1}{NT} C_{NT}(\nu, \nu) \geq \frac{1}{T} C_T(\tilde{\mu}, \tilde{\mu}) \geq \alpha_T \geq h,$$

where $\tilde{\mu} = \frac{1}{N} \sum_{i=1}^N \mu_i$.

Finally, it follows from Theorem 4.3 that $\lim_{N \rightarrow \infty} \frac{1}{NT} C_{NT}(\nu, \nu) = h$. This finishes the proof. \square

Acknowledgements Andrei Agrachev was supported by PRIN and Paul W. Y. Lee was supported by the NSERC postdoctoral fellowship.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Agrachev, A.A., Sachkov, Y.L.: Control Theory from the Geometric Viewpoint. Springer, Berlin (2004)
2. Bernard, P., Buffoni, B.: Optimal mass transportation and Mather theory. *J. Eur. Math. Soc.* **9**(1), 85–121 (2007)
3. Bernard, P., Buffoni, B.: Weak KAM pairs and Monge–Kantorovich duality. Asymptotic analysis and singularities—elliptic and parabolic PDEs and related problems. *Adv. Stud. Pure Math.*, 47-2, pp. 397–420. Mathematical Society of Japan, Tokyo (2007)
4. Brunovsky, P., Lobry, C.: Controllabilité Bang-Bang, controllabilité différentiable, et perturbation des systèmes nonlinéaires. *Ann. Mat. Pura Appl.* **105**, 93–119 (1975)
5. Cannarsa, P., Rifford, L.: Semiconcavity results for optimal control problems admitting no singular minimizing controls. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **25**(4), 773–802 (2008)
6. Cannarsa, P., Sinestrari, C.: *Semiconcave Functions, Hamilton–Jacobi Equations, and Optimal Control*. Birkhäuser, Boston (2004)
7. Evans, L.C.: *Partial Differential Equations*. Graduate Studies in Mathematics, 19. American Mathematical Society, Providence (1998)
8. Fathi, A.: Théorème KAM faible et théorie de Mather sur les systèmes lagrangiens. *C. R. Acad. Sci. Paris Sér. I Math.* **324**(9), 1043–1046 (1997)
9. Fathi, A.: The Weak KAM Theorem in Lagrangian Dynamics, 10th preliminary version
10. Gomes, D.: Hamilton–Jacobi methods for vakonomic mechanics. *NoDEA Nonlinear Differ. Equ. Appl.* **14**(3–4), 233–257 (2007)
11. Kantorovich, L.: On the translocation of masses. *C.R. Doklady Acad. Sci. URSS N.S.* **37**, 199–201 (1942)
12. Lions, P.-L., Papanicolaou, G., Varadhan, S.R.S.: Homogenization of Hamilton–Jacobi equation, unpublished preprint (1987)
13. Montgomery, R.: *A Tour of Subriemannian Geometries, Their Geodesics and Applications*. Mathematical Surveys and Monographs, 91. American Mathematical Society, Providence (2002)
14. Villani, C.: *Topics in Mass Transportation*. American Mathematical Society, Providence (2003)
15. Villani, C.: *Optimal Transport: Old and New*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 338. Springer, Berlin (2009)