

# Optimal Transportation for Generalized Lagrangian

Ji LI<sup>1</sup>     Jianlu ZHANG<sup>1</sup>

**Abstract** This paper deals with the optimal transportation for generalized Lagrangian  $L = L(x, u, t)$ , and considers the following cost function:

$$c(x, y) = \inf_{\substack{x(0)=x \\ x(1)=y \\ u \in \mathcal{U}}} \int_0^1 L(x(s), u(x(s), s), s) ds,$$

where  $\mathcal{U}$  is a control set, and  $x$  satisfies the ordinary equation

$$\dot{x}(s) = f(x(s), u(x(s), s)).$$

It is proved that under the condition that the initial measure  $\mu_0$  is absolutely continuous w.r.t. the Lebesgue measure, the Monge problem has a solution, and the optimal transport map just walks along the characteristic curves of the corresponding Hamilton-Jacobi equation:

$$\begin{cases} V_t(t, x) + \sup_{u \in \mathcal{U}} (V_x(t, x), f(x, u(x(t), t), t) - L(x(t), u(x(t), t), t)) = 0, \\ V(0, x) = \phi_0(x). \end{cases}$$

**Keywords** Optimal control, Hamilton-Jacobi equation, Characteristic curve, Viscosity solution, Optimal transportation, Kantorovich pair, Initial transport measure

**2000 MR Subject Classification** 35D40, 35F21, 37J50

## 1 Introduction

Given a pile of soil and an excavation that we want to fill up with the soil. As early as 1781, Monge posed the question to find an optimal way to do this work. This is the primary statement of Monge problem. We can model the pile of soil and the excavation by two probability measures as the density of the mass. Concretely, given a space  $M$ , and a continuous function, called the cost function  $c(x, y) : M \times M \rightarrow \mathbb{R}$ , given two probability measures  $\mu_0$  and  $\mu_1$  on  $M$ , find the mapping  $t : M \rightarrow M$  which transport  $\mu_0$  to  $\mu_1$  and minimize the total cost  $\int_M c(x, t(x)) d\mu_0$ . But unfortunately, the Monge problem is not always well-posed, for instance, consider  $\mu_0 = \delta_x$ ,  $\mu_1 = \frac{1}{2}(\delta_y + \delta_z)$  ( $x, y, z \in M$ ), the Monge problem has no solution since there is no map  $t$  such that  $t_{\#}\mu_0 = \mu_1$ .

---

Manuscript received February 25, 2014. Revised January 4, 2016.

<sup>1</sup>Department of Mathematics, Nanjing University, 22 Hankou Road, Nanjing 210093, China.

E-mail: adailee.hepburn@gmail.com    zhang@utoronto.ca

It is difficult to study the Monge problem directly. In 1942, Kantorovich raised a generalized problem, later called the Monge-Kantorovich (MK for short) problem, we will state it concretely in Section 4. As proved in [7], MK problem always has a solution. It turned out that through the solution of MK problem, sometimes we can construct the solution of the Monge problem. Recently, several mathematicians studied the connections between Aubry-Mather-Fathi theory, optimal transportation problem and Hamilton-Jacobi equations (see [4-6, 10, 12, 14]).

In [2], they considered a Lagrangian function  $L(x, v, t) : TM \times [0, T] \rightarrow \mathbb{R}$  which satisfies the Tonelli conditions introduced by Mather, and considered the cost by

$$c(x, y) = \min_{\gamma} \int_0^T L(\gamma(t), \dot{\gamma}(t), t) dt,$$

where the minimum is taken on the set of absolutely continuous curve  $\gamma : [0, 1] \rightarrow M$  with  $\gamma(0) = x$  and  $\gamma(1) = y$ . One of the main results of [2] is that using the dual MK problem, they proved that the optimal transformation can be performed by a Borel map with the assumption that the initial measure  $\mu_0$  is absolutely continuous w.r.t. the Lebesgue measure, thus the Monge problem has a solution.

This paper is stimulated by [2], but we use another approach called optimal control, studying the optimal transportation problem of a generalized Lagrangian  $L = L(x, u(x, t), t)$ . Consider the cost function as following:

$$c(x, y) = \inf_{\substack{x(0)=x \\ x(1)=y \\ u \in \mathcal{U}}} \int_0^1 L(x(t), u(x(t), t)) dt,$$

where  $\mathcal{U}$  is called a control set.  $x(t)$  satisfies the following equation:

$$\dot{x} = f(x(t), u(x(t), t)),$$

here  $f$  satisfies some regular condition. About the control set and the regular condition, we will state it in Section 3.

We start from the point of optimal control to study the original optimal transportation problem, called the Monge problem:

$$(M) \quad C(\mu_0, \mu_1) = \min_t \left\{ \int_M c(x, t(x)) d\mu_0 \mid t_{\#} \mu_0 = \mu_1 \right\}.$$

Under the same assumption in [2] that the initial measure  $\mu_0$  is absolutely continuous w.r.t. the Lebesgue measure, we prove that the corresponding Monge problem has a Borel map as its solution.

## 2 Method of Characteristic

In this section, we briefly introduce the method of characteristics in constructing a local classical solutions of the Cauchy problem for Hamilton-Jacobi equation like

$$\begin{cases} \frac{\partial u}{\partial t} + H(x, \nabla u(t, x)) = 0, & (t, x) \in [0, \infty) \times M, \\ u(0, x) = u_0(x), & x \in \mathbb{R}^n, \end{cases} \quad (2.1)$$

where  $H$  and  $u_0$  are of  $C^2$ .

We suppose a priori that we have a solution  $u \in C^2([0, T] \times M)$  of the above equation. We call the solution of the following equation:

$$\begin{cases} \dot{X} = H_p(X, \nabla u(t, X)), \\ X(0) = z, \end{cases} \tag{2.2}$$

$t \rightarrow (t, X(t, z))$ , a characteristic curve associated to  $u$  starting from  $z$ .

Now we set

$$U(t, z) = u(t, X(t, z)), \quad P(t, z) = \nabla u(t, X(t, z)).$$

Easy calculation shows

$$\begin{cases} \dot{U} = -H(X, P) + P \cdot H_p(X, P), \\ \dot{P} = \nabla u_t(t, x) + H_x(x, \nabla u(t, x)) + \nabla^2 u(t, x) H_p(x, \nabla u(t, x)). \end{cases} \tag{2.3}$$

Since

$$0 = \nabla(u_t(t, x) + H(x, \nabla u(t, x))) = \nabla u_t(t, x) + H_x(x, \nabla u(t, x)) + \nabla^2 u(t, x) H_p(x, \nabla u(t, x)),$$

we obtain

$$\dot{P} = -H_x(X, \nabla u(t, X)) = -H_x(X, P).$$

Thus, the pair  $(X, P)$  solves the following ordinary differential equation:

$$\begin{cases} \dot{X} = H_p(X, P), \\ \dot{P} = -H_x(X, P) \end{cases} \tag{2.4}$$

with initial condition  $X(0) = z, P(0) = \nabla u_0(z)$ , while  $U$  satisfies

$$\dot{U} = -H(X, P) + P \cdot H_p(X, P), \quad U(0, z) = u_0(z). \tag{2.5}$$

Obviously,  $X, P$  and  $U$  are uniquely determined by the initial value  $u_0$ . The above arguments suggests that we can obtain a solution to the Hamilton-Jacobi equation (2.1) by solving the characteristic system (2.4) provided the map  $z \rightarrow X(t, z)$  is invertible. In [3], they listed a classic result.

**Theorem 2.1** (see [3]) *For any  $z \in \mathbb{R}^n$ , let  $X(t, z), P(t, z)$  denote the solution of the characteristic system (2.4) and let  $U(t, z)$  be defined by (2.5). Then, there exists  $T^* \geq 0$  such that the map  $z \rightarrow X(t, z)$  is invertible with  $C^1$  inverse  $x \rightarrow Z(t, x)$ , and there exists a unique solution  $u \in C^2([0, T^*) \times M)$  of (2.1), which is given by*

$$u(t, x) = U(t; Z(t; x)), \quad (t, x) \in [0, T^*) \times M. \tag{2.6}$$

**Remark 2.1**  $T^*$  in Theorem (2.1) is the uniform maximum time such that any two characteristic curves do not intersect with each other when the time  $T < T^*$ .

### 3 Optimal Control Problem for Bolza Problem

Optimality is a universal principle of life. The first basic ingredient of an optimal control problem is the so-called control system, and it gives the possible behaviors. Usually, the control system is described by ordinary differential equations. In this paper, we consider the following form:

$$(ODE) \quad \dot{x} = f(x, u), \quad x(t_0) = x_0,$$

where  $x$  is the state,  $x \in M$ ,  $t$  is the time,  $x_0$  is the initial state,  $t_0$  is the initial time,  $u$  is called control which depends on  $x, t$ , i.e.,  $u = u(x, t)$ ,  $t \in \mathbb{R}$ .

Here are some basic assumptions on the control system:

(A1) There exists  $K_0 > 0$  such that  $|f(x, u)| \leq K_0(1 + |x| + |u|)$ ,  $\forall x \in \mathcal{R}^n, u \in \mathcal{U}$ ;

(A2)  $f$  is Lipschitz w.r.t.  $x$ : There exists  $K_1 > 0$  such that  $|f(x_2, u(x_2, t)) - f(x_1, u(x_1, t))| \leq K_1|x_2 - x_1|$  for all  $x_1, x_2 \in M, u \in \mathcal{U}$ ;

(A3)  $f$  is  $C^{1,1}$  with respect to  $x$ : There exists  $K_2 > 0$  such that  $\|f_x(x_2, u(x_2, t)) - f_x(x_1, u(x_1, t))\| \leq K_2|x_2 - x_1|$  for all  $x_1, x_2 \in M, u \in \mathcal{U}$ .

The assumption (A2) ensures the existence of a unique global solution to the control system.

The second basic ingredient is the cost functional, and it associates a cost with each possible behavior. An optimal control problem consists of choosing a control  $u$  in the (ODE) in order to minimize the cost functional.

Let  $L = L(x, u(x, t), t)$  satisfy the following conditions:

(L1)  $L$  is superlinear w.r.t.  $u$ ;

(L2)  $L$  is locally Lipschitz w.r.t.  $x$ :  $\forall R > 0$ , there exists  $\alpha_R$  such that

$$\begin{aligned} |L(x_2, u(x_2, t), t) - L(x_1, u(x_1, t), t)| &\leq \alpha_R|x_2 - x_1|, \quad \forall x_1, x_2 \in B_R, \\ |L(x, u_1(x, t), t) - L(x, u_2(x, t), t)| &\leq C|u_1 - u_2|, \quad \forall x \in M; \end{aligned}$$

(L3)  $\forall x \in M$ , the following set is convex

$$\mathcal{L} = \{(l, v) \in \mathbb{R}^{n+1} \mid \text{there exists } u \in \mathcal{U} \text{ such that } v = f(x, u), l \leq L(x, u)\}.$$

For the convenience and unity, we consider the following optimal control problem:

$$(BP) \quad J(t, x, u) = \int_0^t L(x(s), u(x(s), s))ds + I(x(0)),$$

where  $L \in C^2(M \times \mathbb{R} \times [0, T])$ ,  $I \in C(M)$ , and  $x$  is the solution of the control system:

$$\begin{cases} \dot{x} = f(x(s), u(x(s), s)), \\ x(t) = x. \end{cases} \quad (3.1)$$

Note that fixing endpoint is the same as fixing initial point in essential. Here we use the same symbol  $x$  to denote a point or a solution of (3.1) without confusion. Now, we consider a time period of  $[0, T]$ , and let  $t$  range over  $[0, T]$ ,  $x$  range over  $M$ . We introduce the value function.

**Theorem 3.1** (see [3]) *Assume (A1)–(A2), (L1)–(L3) hold, and the initial cost  $I$  is continuous. Then,  $\forall(x, t) \in M \times [0, T]$ , there exists an optimal control  $u \in \mathcal{U}$  for (BP).*

**Definition 3.1** *Given  $(t, x) \in [0, T] \times M$ , define*

$$V(t, x) = \inf\{J(t, x, u) \mid u : [0, t] \times M \rightarrow \mathcal{U}\},$$

*which is called the value function of the control problem (BP).*

One of the most important principle in optimal control problem is so-called principle of optimality for  $V(t, x)$ .

**Theorem 3.2** (Principle of Optimality) *For every  $(t, x) \in [0, T] \times M$  and  $\Delta t \in (0, t)$ , the value function  $V(t, x)$  satisfies the following relation:*

$$V(t, x) = \inf_{u|_{[t-\Delta t, t]}} \left\{ \int_{t-\Delta t}^t L(x(s), u(x(s), s))ds + V(t - \Delta t, x(t - \Delta t)) \right\}. \quad (3.2)$$

The above principle of optimality means that in order to search for an optimal control, we can search over a small time interval for a control that minimizes the cost over this interval plus the subsequent optimal cost-to-go.

Now, let us look at the value function  $V(t, x)$  in a deep sight. We assume  $V(t, x) \in C^1([0, T] \times M)$ , and consider the right-hand side of (3.2),

$$x(t - \Delta t) = x - f(x, u)\Delta t + o(\Delta t). \quad (3.3)$$

So, we can express  $V(t - \Delta t, x(t - \Delta t))$  as

$$V(t - \Delta t, x(t - \Delta t)) = V(t, x) + V_t(t, x)(-\Delta t) + \langle V_x(t, x), f(x, u)(-\Delta t) \rangle + o(\Delta t), \quad (3.4)$$

and we have

$$\int_{t-\Delta t}^t L(x(s), u(x(s), s))ds = L(x(t), u(x(t), t))\Delta t + o(\Delta t). \quad (3.5)$$

Substituting (3.3) and (3.5) into the right-hand side of (3.2), we get

$$\begin{aligned} V(t, x) = \inf_{u|_{[t-\Delta t, t]}} \{ & L(x(t), u(x(t), t))\Delta t + V(t, x) + V_t(t, x)(-\Delta t) \\ & + \langle V_x(t, x), f(x, u)(-\Delta t) \rangle + o(\Delta t) \}. \end{aligned} \quad (3.6)$$

After simple calculation, we obtain

$$0 = \inf_{u|_{[t-\Delta t, t]}} \{L(x(t), u(x(t), t)) - V_t(t, x) - \langle V_x(t, x), f(x, u) \rangle\}. \quad (3.7)$$

This is equivalent to the following equation:

$$V_t(t, x) + \sup\{\langle V_x(t, x), f(x, u) \rangle - L(x(t), u(x(t), t))\} = 0. \quad (3.8)$$

It is remarkable to note that there are several forms which are equivalent to (3.8), but (3.8) is the reasonable form since flipping the sign in a PDE affects its viscosity solutions, so it will turn out that the above sign convention is the correct one. Let

$$H = \sup_{u \in \mathcal{U}} \langle p, f(x, u) \rangle - L(x, u, t). \tag{3.9}$$

We have the following theorem.

**Theorem 3.3** *If the initial condition is equal to the initial cost, i.e.,  $u_0(x) = I(x)$ , then  $V(t, x)$  is the unique viscosity solution of (2.1) with Hamiltonian function  $H$  as in (3.9).*

**Proof Step 1** (Viscosity subsolution) For every  $C^1$  test function  $\varphi = \varphi(x, t)$ , assume that  $\varphi \geq V$ , and that  $\varphi - V$  attains a local minimum at  $(t_0, x_0)$ . We need to prove

$$\varphi(t, x) + \sup_{u \in \mathcal{U}} \{ \langle \varphi_x(t, x), f(x, u) \rangle - L \} \leq 0.$$

Suppose the contrary: There exists a  $C^1$  function  $\varphi$ , and a control  $u_0$ , s.t.

$$\varphi(t_0, x_0) = V(t_0, x_0), \quad \varphi(t, x) = V(t, x), \quad \forall (t, x) \in U_\delta(t_0, x_0)$$

and

$$\varphi_t(t - 0, x_0) + \langle \varphi_x(t_0, x_0), f(x, u) \rangle - L > 0.$$

Since

$$\begin{aligned} V(t_0 - \Delta t, x(t_0 - \Delta t)) - V(t_0, x_0) &\leq \varphi(t_0 - \Delta t, x(t_0 - \Delta t)) - \varphi(t_0, x_0) \\ &= - \int_{t_0 - \Delta t}^{t_0} \frac{d\varphi(t, x)}{dt} \\ &= - \int_{t_0 - \Delta t}^{t_0} \varphi_t(t, x) + \langle \varphi_x(t, x), f(x, u) \rangle dt \\ &\leq - \int_{t_0 - \Delta t}^{t_0} L dt. \end{aligned} \tag{3.10}$$

This is contradictory to the principle of optimal, so  $V(t, x)$  is a viscosity subsolution.

**Step 2** (Viscosity supersolution) For every  $C^1$  test function  $\varphi = \varphi(t, x)$ , assume that  $\varphi \leq V$ , and that  $\varphi - V$  attains a local minimal at  $(t_0, x_0)$ . We need to prove

$$\varphi(t, x) + \sup_{u \in \mathcal{U}} \{ \langle \varphi_x(t, x), f(x, u) \rangle - L \} \geq 0.$$

Suppose the contrary: There exists a  $C^1$  function  $\psi$ , and a control  $u'_0$ , s.t.

$$\psi(t_0, x_0) = V(t_0, x_0), \quad \psi(t, x) \leq V(t, x), \quad \forall (t, x) \in U_\delta(t_0, x_0)$$

and

$$\psi_t + \langle \psi_x, f \rangle - L < 0.$$

Since

$$\begin{aligned}
 \int_{t_0-\Delta t}^{t_0} L &= V(t_0, x_0) - V(t_0 - \Delta t, x(t_0 - \Delta t)) \\
 &\leq \psi(t_0, x_0) - \psi(t_0 - \Delta t, x(t_0 - \Delta t)) \\
 &= \int_{t_0-\Delta t}^{t_0} \frac{d\psi(t, x)}{dt} = \int_{t_0-\Delta t}^{t_0} \psi_t(t, x) + \langle \psi_x, f(x, u) \rangle dt \\
 &< \int_{t_0-\Delta t}^{t_0} L(x, u) dt.
 \end{aligned} \tag{3.11}$$

We get a contradictory, thus  $V(t, x)$  is a viscosity supersolution.

**Step 3 (Uniqueness)** By the comparison principle, the viscosity solution is unique!

From the above three steps, we have proved that  $V(t, x)$  is the unique viscosity solution.

Let the Hamiltonian  $H$  be as (3.9). We have the following theorem.

**Theorem 3.4** (see [3]) *Let  $(u, x)$  be an optimal pair for the point  $(t_0, x_0) \in [0, T] \times M$  and let  $p : [0, t] \rightarrow \mathbb{R}^n$  be a dual arc associated with  $(u, x)$ . Then  $(x, p)$  solves the system*

$$\begin{cases} \dot{x}(s) = H_x(x(s), p(s)), \\ \dot{p}(s) = -H_p(x(s), p(s)) \end{cases} \tag{3.12}$$

for all  $s \in [0, t]$ . As a consequence,  $x, p$  are of class  $C^1$ .

**Remark 3.1** The above theorem tells us that the optimal trajectory of the optimal control problem is nothing else, but just the characteristic curve of the corresponding Hamilton-Jacobi equation.

### 4 Optimal Transportation Related to $L(x, u, t)$

**Statement of the Monge problem** Let  $M$  be a compact manifold, and a continuous cost function  $c(x, y)$  is given,  $c(x, y) : M \times M \rightarrow \mathbb{R}$ ,  $\mu_0, \mu_1$  are two probability measures on  $M$ . Find the mapping  $\Psi : M \rightarrow M$  which transport  $\mu_0$  into  $\mu_1$ , and minimize the total cost:

$$\int_M c(x, \Psi(x)) d\mu_0,$$

where the transport means push-forward, which is defined as following.

**Definition 4.1** (Push-forward) *Let  $t : X \rightarrow Y$  be a measurable map,  $\mu$  is a measure. Define*

$$t_{\#}\mu(B) = \mu(t^{-1}(B))$$

for every Borel set  $B \subset Y$ .

The Monge problem can be stated as

$$(M) := \min \left\{ \int_{X \times Y} c(x, \Psi(x)) d\mu_0 \mid \Psi : M \rightarrow M, \Psi_{\#}\mu_0 = \mu_1 \right\}.$$

In the rest of this paper, we consider the cost function as following:

$$c(x, y) = \inf_{\substack{x(0)=x \\ x(1)=y \\ u \in \mathcal{U}}} \int_0^1 L(x(s), u(x(s), s)) ds, \tag{4.1}$$

where  $x$  satisfies the following ordinary equation:

$$\dot{x}(s) = f(x(s), u(x(s), s)). \tag{4.2}$$

As we said in the introduction, the Monge problem is not always well-posed, since sometimes there is no map  $\Psi$  such that  $\Psi_{\#}\mu_0 = \mu_1$ . In the study of the Monge problem, mathematicians find that it is a good and effective approach to consider the MK problem first.

**Statement of the Monge-Kantorovich problem**

$$(\text{MK}) := \min \left\{ \int_{X \times Y} c(x, y) d\gamma(x, y) \mid \gamma \in \mathcal{P}(X \times Y), (\pi_1)_{\#}\gamma = \mu_0, (\pi_2)_{\#}\gamma = \mu_1 \right\}.$$

Under some proper conditions, we can use the solution of the MK problem to construct the solution of the Monge problem. In this paper, we use this approach.

The measure  $\gamma$  which satisfies  $(\pi_1)_{\#}\gamma = \mu_0, (\pi_2)_{\#}\gamma = \mu_1$  in (MK) is called a transport plan. We denote the transport plans by  $\mathcal{K}(\mu_0, \mu_1)$ . Observe that if  $\Psi$  is admissible for Monge problem, i.e.,  $\Psi_{\#}\mu_0 = \mu_1$ , then the measure  $\gamma = (Id \times \Psi)_{\#}\mu_0$  is a transport plan for (MK). Moreover, the class of transport plans is always non-empty since it always contains  $\mu_0 \times \mu_1$ . As proved in [7], the min in (MK) always can be achieved by a transport plan, and we call the transport plan which realizes the min an optimal transfer plan (see [2]).

Let the total cost

$$C(\mu_0, \mu_1) = \min_{\gamma \in \mathcal{K}(\mu_0, \mu_1)} \int c(x, y) d\gamma. \tag{4.3}$$

**Definition 4.2** (Admissible Kantorovich Pair) *A pair of continuous function  $(\phi_0, \phi_1)$  is called an admissible Kantorovich pair if it satisfies the following relations:*

$$\phi_1(x) = \min_{y \in M} \phi_0(y) + c(x, y), \quad \phi_0(x) = \max_{y \in M} \phi_1(y) - c(x, y), \tag{4.4}$$

for all point  $x \in M$ .

Kantorovich proved the following notable result.

**Theorem 4.1**

$$C(\mu_0, \mu_1) = \max_{(\phi_0, \phi_1)} \left( \int_M \phi_1 d\mu_1 - \int_M \phi_0 d\mu_0 \right), \tag{4.5}$$

where the max is taken on the set of admissible Kantorovich pair  $(\phi_0, \phi_1)$ .

This is the so-called dual Kantorovich problem. The admissible pairs which attain the max are called optimal Kantorovich pairs, usually, it is not unique!



**Lemma 4.1** *If  $\gamma$  is an optimal transfer plan, and  $(\phi_0, \phi_1)$  is an optimal Kantorovich pair, then the support of  $\gamma$  is contained in the following set:*

$$\{(x, y) \in M \times M \text{ such that } \phi_1(y) - \phi_0(x) = c(x, y)\} \subset M \times M.$$

**Lemma 4.2** *We have*

$$c(x, y) = \max_{(\phi_0, \phi_1)} \phi_1(y) - \phi_0(x),$$

where the max is taken on the set of admissible Kantorovich pairs.

It is deserved to note that for different pair of  $(x, y)$ , the max may be achieved by different admissible Kantorovich pair.

Let  $(\phi_0, \phi_1)$  be an optimal Kantorovich pair, and

$$c_0^t(x, y) = \inf_{\substack{x(0)=x \\ x(1)=y \\ u \in \mathcal{U}}} \int_0^t L(x(s), u(x(s), s)) ds.$$

We construct a function on  $M \times [0, T]$  as following:

$$U(x, t) = \min_{y \in M} \phi_0(x) + c_0^t(x, y). \tag{4.6}$$

Recall the definition of the value function for (BP):

$$V(t, x) = \inf_{u \in \mathcal{U}} \int_0^t L(x(s), u(x(s), s), s) ds + I(x(0)),$$

where  $x$  satisfies the control system (3.1). Obviously, we can see easily that  $U(x, t)$  is just the same as the value function  $V(t, x)$  with  $I(x) = \phi_0(x)$ .

So, the total cost  $C(\mu_0, \mu_1)$  can be realized by a viscosity solution of the Hamilton-Jacobi equation (2.1) with a proper initial condition. We have

$$C(\mu_0, \mu_1) = \int_M V(1, x) d\mu_1 - \int_M V(0, x) d\mu_0. \tag{4.7}$$

Let us stop here for a little while, and look at the value function from another point of view. Assume that  $u^*$  realizes the inf in the definition of value function  $V(t, x)$ . We define an operator  $\mathcal{T} : C^{ac}([0, t], R) \rightarrow C^{ac}([0, t], R)$  as following:

$$\mathcal{T}_t \phi_0(x) = \phi_0(\gamma(0)) + \int_0^t L(\gamma(s), u^*(\gamma(s), s), s) ds. \tag{4.8}$$

Obviously,  $\mathcal{T}_t \phi_0(x) = V(t, x)$ .

As in [11, 13], we call  $\mathcal{T}_t$  the solution semigroup since it is determined by the viscosity solution of the Hamilton-Jacobi equation (2.1). Actually,  $\mathcal{T}_t$  is indeed a semigroup.

**Theorem 4.2** (Semi-group Property)  *$\{\mathcal{T}_t\}_{t \geq 0}$  is a one-parameter semigroup from  $C(M, \mathbb{R})$ .*

**Proof**  $\mathcal{T}_s\phi_0(x) = u(x, s)$  means that the operator  $\mathcal{T}_s$  just sends the initial value  $\phi_0$  into the corresponding viscosity solution at time  $s$  under this initial condition. Similarly,  $\mathcal{T}_{t+s}\phi_0(x) = u(x, s + t)$  means that the operator  $\mathcal{T}_{t+s}$  sends the initial value  $\phi_0$  into the corresponding viscosity solution at time  $t + s$ . Attention  $\mathcal{T}_t \circ \mathcal{T}_s\phi_0(x)$  means that the operator  $\mathcal{T}_t$  sends the initial value  $\mathcal{T}_s\phi_0$  to the corresponding viscosity solution at time  $t$ . Since the viscosity solution of Hamilton-Jacobi (2.1) is unique, it is obvious that  $\mathcal{T}_t \circ \mathcal{T}_s\phi_0 = \mathcal{T}_{t+s}\phi_0$  for arbitrary initial value  $\phi_0 \in C(M, \mathbb{R})$ .

**Definition 4.3** Let  $(\phi_0, \phi_1)$  denote an optimal Kantorovich pair, and  $u^*$  be one of the control which reaches the inf in the definition of the value function  $V$ . We define  $\mathcal{F}(\phi_0, \phi_1) \subset C^2([0, 1], M)$  as the set of curves  $\gamma(t)$  such that

$$\phi_1(\gamma(1)) = \phi_0(\gamma(0)) + \int_0^1 L(\gamma(t), u^*(\gamma(t), t))dt. \tag{4.9}$$

**Remark 4.1** Obviously, if we rewrite (4.9) into the following form:

$$V(1, \gamma(1)) = V(0, \gamma(0)) + \int_0^1 L(\gamma(t), u^*(t, \gamma(t), t))dt, \tag{4.10}$$

we can see that  $\mathcal{F}(\phi_0, \phi_1)$  is just the set of pieces of characteristic curves.

Let  $\mathcal{F}_0(\psi_0, \psi_1)$  be the set of initial state  $(x, p) \in T^*M$  such that the curve  $t \rightarrow \pi \circ \psi_0^t(x, v)$  belongs to  $\mathcal{F}(\phi_0, \phi_1)$ . We have the following lemma.

**Lemma 4.3** The maps  $\pi$  and  $\pi \circ \psi_0^1 : \mathcal{F}_0(\phi_0, \phi_1) \rightarrow M$  are surjective. If  $x$  is a differentiable point of  $\phi_0$ , then the set  $\pi^{-1}(x) \cap \mathcal{F}(\phi_0, \phi_1)$  contains only one point. There exists a Borel measurable set  $\Sigma \subset M$  of full measure, whose points are differentiable points of  $\phi_0$ , and such that the map

$$x \rightarrow S(x) = \pi^{-1}(x) \cap \mathcal{F}(\phi_0, \phi)$$

is Borel measurable on  $\Sigma$ .

**Proof** For each  $x \in M$ , there exists a characteristic curve such that (4.9)–(4.10) are satisfied, so the projection  $\pi \circ \psi_0^1$  from  $\mathcal{F}_0(\phi_0, \phi_1)$  to  $M$  is surjective. For  $\pi$ , it is similar!

Next we consider a differentiable point  $x$  of  $\phi_0$ . By the characteristic method introduced in Section 2, the characteristic curves do not cumulate together at  $x$ , i.e., there is only one characteristic curve starts from  $x$ . We construct  $S$  as

$$S(x) = \pi^{-1}(x) \cap \mathcal{F}(\phi_0, \phi). \tag{4.11}$$

Since  $\phi_0$  is Lipschitz, the set of differentiable points of  $\phi_0$  is of total Lebesgue measure. Notice that there exists a sequence of compact sets  $K_n$  such that  $\phi_0$  is differentiable at each point of  $K_n$ , and the Lebesgue measure of  $M - K_n$  is converging to zero. For each  $n$ , the set  $\pi^{-1}(K_n) \cap \mathcal{F}(\phi_0, \phi_1)$  is compact, and the canonocal projection  $\pi$  restricted to this set is injective and continuous, so the inverse function  $S$  is continuous on  $K_n$ . And as a consequence,  $S$  is a Borel measurable map on  $\Sigma = \bigcup_n K_n$ .

Now let

$$H(x, p) = \max_u \langle p, f(x, u) \rangle - L(x, u, t).$$

And let  $m_0 \in \mathcal{B}(T^*M)$  be a Borel probability measure on the cotangent bundle  $T^*M$ . We call  $m_0$  an initial transport measure if the measure  $\eta$  on  $M \times M$  given by

$$\eta = (\pi \times (\pi \circ \psi))_{\#} m_0$$

is a transport plan. Here  $\psi$  is the time one map of the Hamiltonian flow, and  $\pi : T^*M \rightarrow M$  is the canonical projection. We denote  $\mathcal{I}(\mu_0, \mu_1)$  the set of initial transport measures. Obviously, we can define the action of an initial transport measure as

$$A(m_0) = \int_{M \times M} c(x, y) d\eta. \tag{4.12}$$

**Lemma 4.4** *The mapping  $(\pi \times (\pi \circ \psi))_{\#} : \mathcal{I}(\mu_0, \mu_1) \rightarrow \mathcal{K}(\mu_0, \mu_1)$  is surjective.*

**Proof** We shall prove that for arbitrary probability measure  $\eta \in \mathcal{B}(M \times M)$ , there exists a probability measure  $m_0 \in \mathcal{B}(T^*M)$  such that  $(\pi \times (\pi \circ \psi_0^T))_{\#} m_0 = \eta$ . Since the set of probability measures on  $M \times M$  is the compact convex closure of the set of Dirac probability measures, we just need to prove it when  $\eta$  is a Dirac probability measure. Let  $\eta$  be a Dirac probability measure supported at  $(x, y) \in M \times M$ , and  $\gamma : [0, T] \rightarrow M$  be a curve with boundary conditions  $\gamma(0) = x$ ,  $\gamma(1) = y$  satisfying the control system. Let  $m_0$  be a Dirac probability measure supported at  $\gamma(0), \dot{\gamma}(0)$ . Obviously, we have

$$\eta = (\pi \times (\pi \circ \psi_0^T))_{\#} m_0,$$

where  $\eta$  is a transport plan which is determined by  $m_0$  through  $(\pi \times (\pi \circ \psi))_{\#}$ .

**Theorem 4.3** *If  $\mu_0$  is absolutely continuous with respect to Lebesgue measure, then, there exists a Borel section  $S : M \rightarrow TM$  such that the map  $\pi \circ \psi_0^1 \circ S$  is an optimal transport map between  $\mu_0$  and  $\mu_1$ . The section  $S$  is unique in the sense of  $\mu_0$  almost everywhere.*

**Proof** Since  $\Sigma$  is of full Lebesgue measure, consider  $m_0 = S_{\#}(\mu_0|_{\Sigma})$ , which is a probability measure on  $T^*M$ , concentrated on  $\mathcal{F}(\phi_0, \phi_1)$ , and  $\pi_{\#} m_0 = \mu_0$ . Since  $\pi$  induces an isomorphism between  $\pi^{-1}(\Sigma) \cap \mathcal{F}_0(\phi_0, \phi_1)$ , and by Lemma 4.4, we have  $C_0^1(\mu_0, \mu_T) = \min_{\mathcal{I}} A(m_0)$ , and that  $m_0$  is the unique initial transport measure. Thus,  $m_0$  is the unique optimal transport measure.  $\pi \circ \psi \circ S$  is an optimal transport map between  $\mu_0$  and  $\mu_1$ .

**Acknowledgement** We appreciate the anonymous referees for their valuable comments.

## References

- [1] Ambrosio, L., Lecture Notes on Optimal Transport Problems Mathematical Aspects of Evolving Interfaces, Lecture Notes in Mathematics, Vol. 1812, 2004, 1–52.
- [2] Bernard, P. and Buffoni, B., Optimal mass transportation and Mather theory, *J. Eur. Math. Soc.*, **9**, 2007, 85–121.
- [3] Cannarsa, P. and Sinestrari, C., Semiconcave Functions, Hamilton-Jacobi Equations, and Optimal Control Progress in Nonlinear Differential Equations and Their Applications, 2004.

- [4] De Passcale, L., Gelli, M. S. and Granieri, L., Minimal measures, one-dimensional currents and the Monge-Kantorovich problem, *Calculus of Variations and Partial Differential Equations*, **27**(1), 2006, 1–23.
- [5] Evans, L. C. and Gomes, D., Linear programming interpretations of Mather’s variational principle, Attribute to J. L. Lions, *Esaim Control Optim, Calc. Var.*, **8**, 2002, 693–702.
- [6] Granieri, L., On action minimizing measures for the Monge-Kantorovich problem, *NoDEA*, **14**(1–2), 2007, 125–152.
- [7] Kantorovich, L. V., On the transfer of mass, *Dokl. Akad. Nauk. USSR*, **37**, 1942, 227–229.
- [8] Kantorovich, L. V., On a problem of Monge, *Uspekhi Mat. Nauk.*, **3**, 1948, 225–226.
- [9] Liberzon, D., *Calculus of Variations and Optimal Control Theory: A Concise Introduction*, ISBN: 97814008426431.
- [10] Pratelli, A., Equivalence between some definitions for the optimal mass transportation problem and for the transport density on manifolds, *Annali di Math. Pura App.*, 2004.
- [11] Su, X. F. and Yan, J., Weak KAM theorem for Hamilton-Jacobi equations, preprint. arXiv: 1312.1600
- [12] Villani, C., *Topics in Optimal Transportation*, American Mathematical Society, Providence, Rhode Island, 2003.
- [13] Wang, L. and Yan, J., Uniqueness of viscosity solutions of Hamilton-Jacobi equations, Weak KAM theory for general Hamilton-Jacobi equations II: The fundamental solution under Lipschitz conditions. arXiv: 1408.3791v1
- [14] Wolansky, G., Optimal transportation in the presence of a prescribed pressure field. arXiv: math-ph10306070
- [15] Young, L. C., *Lectures on the Calculus of Variations and Optimal Control Theory*, 2nd edition, Chelsea, 1980, *Econometrica*, **39**(3), 1971, 653.