# Limit Theorems for Optimal Mass Transportation

G. Wolansky[1]
Department of Mathematics,
Technion, Haifa 32000, Israel

**Abstract**

The optimal mass transportation was introduced by Monge some 200 years ago and is, today, the source of large number of results in analysis, geometry and convexity. Here I investigate a new, surprising link between optimal transformations obtained by different Lagrangian actions on Riemannian manifolds. As a special case, for any pair of non-negative measures $\lambda^+, \lambda^-$ of equal mass

$$W_1(\lambda^-, \lambda^+) = \lim_{\varepsilon\to 0} \varepsilon^{-1} \inf_{\mu} W_p(\mu + \varepsilon\lambda^-, \mu + \varepsilon\lambda^+)$$

where $W_p$, $p \geq 1$ is the Wasserstein distance and the infimum is over the set of probability measures in the ambient space.

## 1 Introduction

The Wasserstein metric $W_p$ ($\infty > p \geq 1$) is a useful distance on the set of positive Borel measures on metric spaces. Given a metric space $(M, D)$ and a pair of positive Borel measures $\lambda^\pm$ on $M$ satisfying $\int_M d\lambda^+ = \int_M d\lambda^-$:

$$W_p(\lambda^+, \lambda^-) := \inf_{\pi} \left\{ \left[ \int_M \int_M D^p(x, y) d\pi(x, y) \right]^{1/p} ; \quad \pi \in \mathcal{P}(\lambda^+, \lambda^-) \right\} , \tag{1.1}$$

where $\mathcal{P}(\lambda^+, \lambda^-)$ stands for the set of all positive Borel measures on $M \times M$ whose $M-$marginals are $\lambda^+, \lambda^-$.

Under fairly general conditions (e.g if $M$ is compact), a minimizer $\pi^0 \in \mathcal{P}(\lambda^+, \lambda^-)$ of (1.1) exists. Such minimizers are called *optimal plans.* I'll assume in this paper that $M$ is a compact Riemannian manifold and $D$ is a metric related (but not necessarily identical) to the geodesic distance.

If in addition $\lambda^+$ satisfies certain regularity conditions, the optimal measure $\pi^0$ is supported on a graph of a Borel mapping $\Psi : M \to M$. By some abuse of notation we call a Borel map $\Psi$ *an optimal plan* if it is a minimizer of

$$W_p(\lambda^+, \lambda^-) = \inf_{\Phi} \left\{ \left[ \int D^p(x, \Phi(x)) d\lambda^+ \right]^{1/p} ; \quad \Phi_\# \lambda^+ = \lambda^- \right\}$$

(see Section 1.2-4 for notation).

[1]This research was supported by M. & M. Bank Mathematics Res. Fund and by the Israel Science Foundation

The metric $W_p$, $p \geq 1$ is a metrization of the weak topology $C^*(M)$ on positive Borel measures. In particular, it is continuous in the weak topology. Thus, it is possible to approximate $W_p(\lambda^+, \lambda^-)$ (and the corresponding optimal plan) by $W_p(\lambda^+_N, \lambda^-_N)$ on the set of *atomic measures*

$$\lambda^\pm_N \in \mathcal{M}^{+,N} := \left\{ \mu = \sum_{i=1}^N m_i \delta_{(x_i)} \quad , m_i \geq 0, \quad x_i \in M \right\} \ , \ N \to \infty \tag{1.2}$$

reducing (1.1) into a *finite-dimensional linear programming* on the set of non-negative $N \times N$ matrices $\{\mathcal{P}_{i,j}\}$ subjected to linear constraints.

There is, however, a sharp distinction between the case $p > 1$ and $p = 1$. If $p > 1$ then the optimal plan $\pi^0$ is unique (for regular $\lambda^+$). This is, in general, *not the case* for $p = 1$. Another distinctive feature of the case $p = 1$ is its "pinning property": The distance $W_1$ depends only on the difference $\lambda := \lambda^+ - \lambda^-$. This is manifested by the alternative, dual formulation of $W_1$:

$$W_1(\lambda) = \sup_\phi \left\{ \int \phi d\lambda \ ; \ \|\phi\|_{Lip} \leq 1 \right\} \tag{1.3}$$

where $\|\phi\|_{Lip} := \sup_{x \neq y \in M} \left(\phi(x) - \phi(y)\right) / D(x, y)$.
The optimal potential $\phi$ yields some partial information on the optimal plan $\Psi$ (if exists). In particular, $\nabla\phi(x)$, whenever exists, only indicates *the direction* of the optimal plan. For example, if the metric $D$ is Euclidean, then $\Psi(x) = x + t(x)\nabla\phi(x)$ for some unknown $t(x) \in \mathbb{R}^+$. This is in contrast to the case $p > 1$ where a dual variational formulation, analogous to (1.3), yields the *complete information* on the optimal plan $\Psi$ in terms of the gradient of some potential $\phi$.

In this paper I consider an object called the $p-$Wasserstein distance ($p > 1$) of $\lambda^+$ to $\lambda^-$, *conditioned on a probability measure* $\mu$:

$$W^{(p)}(\lambda \| \mu) := \sup_\phi \left\{ \int \phi d\lambda \ ; \ \int |\nabla\phi|^q d\mu \leq 1 \right\} \tag{1.4}$$

where $q = p/(p-1)$.

The first result is

$$W_1(\lambda) = \min_\mu \left\{ W^{(p)}(\lambda \| \mu) \ ; \ \int d\mu = 1 \right\} \quad , \ (p > 1) \tag{1.5}$$

The problem associated with (1.5) is related to *shape optimization*, see [7]. In addition, the minimizer $\mu$ in (1.5) and the corresponding maximizer $\phi$ in (1.4) or (1.3) play an important rule in the $L_1$ theory of transport [12]. In fact, the optimal $\phi$ is, in general, a Lipschitz function which is differentiable $\mu$ a.e. and satisfies $|\nabla\phi| = 1$ $\mu$ a.e. The minimal measure $\mu$ is called a *transport measure*. It verifies the weak form of the continuity equation which, under the current notation, takes the form

$$\nabla \cdot (\mu \nabla\phi) = \frac{\lambda}{W_1(\lambda)} \ .$$

The transport measure yields an additional information on the optimal plan $\Psi$ along the *transport rays* which completes the information included in $\nabla\phi$ [12]. In the context of shape optimization it is related to the optimal distribution of conducting material [7]. See also [19], [23], [24].

The evaluation of the transport measure $\mu$ is therefore an important object of study. It is tempting to approximate the transport measure as a minimizer of (1.5) on a restricted finite space, e.g. for $\mu \in \mathcal{M}^{+,N}$ as defined in (1.2).

However, this cannot be done. Unlike $W_p$, $W^{(p)}(\lambda\|\mu)$ is *not* continuous in the weak topology of $C^*$ on Borel measures with respect to both $\mu$ and $\lambda$. Indeed, it follows easily that $W^{(p)}(\lambda\|\mu) = \infty$ for any atomic measure $\mu$.

The second result of this paper is

$$W^{(p)}(\lambda\|\mu) = \lim_{n\to\infty} nW_p(\mu + \lambda^+/n, \mu + \lambda^-/n) \tag{1.6}$$

Here the limit is in the sense of $\Gamma$ convergence. A somewhat stronger result is obtained if we take the infimum over all probability measures $\mu$:

$$W_1(\lambda) = \lim_{n\to\infty} n \min_{\mu} W_p(\mu + \lambda^+/n, \mu + \lambda^-/n) \tag{1.7}$$

where the convergence is, this time, pointwise in $\lambda$.

The importance of (1.6, 1.7) is that $W^{(p)}(\lambda\|\mu)$ can now be approximated by a *weakly continuous function*

$$W_n^{(p)}(\lambda^+, \lambda^-\|\mu) := nW_p(\mu + \lambda^+/n, \mu + \lambda^-/n)\ .$$

Suppose $\mu_0$ is a unique minimizer of (1.5). If $\mu_n$ is a minimizer of $W_n^{(p)}(\lambda^+, \lambda^-\|\mu)$ then the sequence $\{\mu_n\}$ must converge to the transport measure $\mu_0$. In contrast to $W^{(p)}$, $W_n^{(p)}$ *is continuous* in the $C^*$ topology with respect to $\mu$. Hence $\mu_n$ can be approximated by atomic measures $\mu_n^N \in \mathcal{M}^{+,N}$ (1.2). In particular a transport measure can be approximated by a finite *points allocation* obtained by minimizing $W_n^{(p)}$ on $\mathcal{M}^{+,N}$ for a sufficiently large $n$ and $N$.

The results (1.5- 1.7) can be extended to the case where the cost $D^p$ on $M \times M$ is generalized into an action function on a Riemannian manifold $M \times M$, induced by a Lagrangian function $l : TM \to \mathbb{R}$. This point of view reveals some relations with the *Weak KAM Theory* dealing with invariant measures of Lagrangian flows on manifolds.

## 1.1 Overview

Section 2 review the necessary background for the Weak KAM and its relation to optimal transport. Section 3 state the main results (Theorems 1-4), which correspond to (1.5- 1.7) for homogeneous Lagrangian on $M \times M$. Section 4 presents the proof of the first of the main results which generalizes (1.4). Finally, Section 5 contains the proofs of the other main results which generalize (1.6, 1.7).

## 1.2 Standing notations and assumptions

1. $(M, g)$ is a compact, Riemannian Manifold and $D : M \times M \to \mathbb{R}^+$ is the geodesic distance.

2. $TM$ (res. $T^*M$) the tangent (res. cotangent) bundle of $M$. The duality between $v \in T_xM$ and $p \in T_x^*M$ is denoted by $\langle \xi, v\rangle \in \mathbb{R}$. The projection $\Pi : TM \to M$ is the trivialization $\Pi(x, v) = x$. Likewise $\Pi^* : T^*M \to M$ is the trivialization $\Pi^*(x, \xi) = x$.

3. For any topological space $X$, $\mathcal{M}(X)$ is the set of Borel measures on $X$, $\mathcal{M}_0(X) \subset \mathcal{M}(X)$ the set of such measures which are perpendicular to the constants. $\mathcal{M}^+(X) \subset \mathcal{M}(X)$ the set of all non-negative measures in $\mathcal{M}$, and $\mathcal{M}_1^+(X) \subset \mathcal{M}^+(X)$ the set of normalized (probability) measures. If $X = M$, the parameter $X$ is usually omitted.

4. A Borel map $\Phi : X_1 \to X_2$ induces a mapping $\Phi_\# : \mathcal{M}^+(X_1) \to \mathcal{M}^+(X_2)$ via
$$\Phi_\#(\mu_1)(A) = \mu_1(\Phi^{-1}(A))$$
for any Borel set $A \subset X_2$.

5. For any $x, y \in M$ let $\mathcal{K}_{x,y}^T$ be the set of all absolutely continuous paths $\boldsymbol{z} : [0, T] \to M$ connecting $x$ to $y$, that is, $\boldsymbol{z}(0) = x$, $\boldsymbol{z}(T) = y$.

6. Given $\mu_1, \mu_2 \in \mathcal{M}^+$, the set $\mathcal{P}(\mu_1, \mu_2)$ is defined as all the measures $\Lambda \in \mathcal{M}^+(M \times M)$ such that $\pi_{1,\#}\Lambda = \mu_1$ and $\pi_{2,\#}\Lambda = \mu_2$, where $\pi_i : \mathcal{M} \times \mathcal{M} \to \mathcal{M}$ defined by $\pi_1(x, y) = x$, $\pi_2(x, y) = y$.

7. An hamiltonian function $h \in C^2(T^*M; \mathbb{R})$ is assumed to be strictly convex and super-linear in $\xi$ on the fibers $T_x^*M$, uniformly in $x \in M$, that is
$$h(x, \xi) \geq -C + \hat{h}(\xi) \ \text{ where } \lim_{\|\xi\| \to \infty} \hat{h}(\xi)/\|\xi\| = \infty \quad .$$
The Lagrangian $l : TM \to \mathbb{R}$ is obtained by Legendre duality
$$l(x, v) = \sup_{\xi \in T_x^*M} \langle \xi, v\rangle - h(x, \xi)$$
satisfies $l \in C^2(TM; \mathbb{R})$, and is super linear on the fibers of $T_xM$ uniformly in $x$.

8. $Exp_{(l)} : TM \times \mathbb{R} \to M$ is the flow due to the Lagrangian $l$ on $M$, corresponding to the Euler-Lagrange equation
$$\frac{d}{dt} l_v = l_x \ .$$
For each $t \in \mathbb{R}$, $Exp_{(l)}^{(t)} : TM \to M$ is the exponential map at time $t$.

## 2 Background

The weak version of Mather's theory [20] deals with minimal invariant measures of Lagrangians, and the corresponding Hamiltonians defined on a manifold $M$. In this theory the concept of an orbit $\boldsymbol{z} = \boldsymbol{z}(t) : \mathbb{R} \to M$ is replaced by that of a *closed probability measure* on $TM$:

$$\mathcal{M}_0^c := \left\{ \nu \in \mathcal{M}_1^+(TM) \ ; \ \int_{TM} l(x,v) d\nu(x,v) < \infty \ , \ \int_{TM} \langle d\phi, v\rangle d\nu = 0 \ \text{ for any } \phi \in C^1(M) \right\} . \tag{2.1}$$

A minimal (or Mather) measure $\nu_M \in \mathcal{M}_0^c$ is a minimizer of

$$\inf_{\nu \in \mathcal{M}_0^c} \int_{TM} l(x,v) d\nu(x,v) := -\underline{E} \tag{2.2}$$

It can be shown ([2], [18], [3]) that any minimizer of (2.2) is invariant under the flow induced by the Euler-Lagrange equation on $TM$:

$$\frac{d}{dt} \nabla_{\dot{x}} l(x, \dot{x}) = \nabla_x l(x, \dot{x}) \ . \tag{2.3}$$

There is also a dual formulation of (2.2) [17], [29]:

$$\sup_{\mu \in \mathcal{M}_1^+} \inf_{\phi \in C^1(M)} \int_M h(x, d\phi) d\mu = \underline{E} \ , \tag{2.4}$$

where the maximizer $\mu_M$ is the projection of a Mather measure $\nu_M$ on $M$. The ground energy level $\underline{E}$, common to (2.2, 2.4), admits several equivalent definitions. Evans and Gomes ([11] [13] [14]) defined $\underline{E}$ as the *effective hamiltonian value*

$$\underline{E} := \inf_{\phi \in C^1(M)} \sup_{x \in M} h(x, d\phi) \ ,$$

while the PDE approach to the WKAM theory ([16], [17]) defines $\underline{E}$ as the minimal $E \in \mathbb{R}$ for which the Hamilton-Jacobi equation $h(x, d\phi) = E$ admits a viscosity sub-solution on $M$. Alternatively $\underline{E}$ is the *only* constant for which $h(x, d\phi) = \underline{E}$ admits a viscosity solution [15]. There are other, equivalent definitions of $\underline{E}$ known in the literature. We shall meet some of them below.

**Example 2.1.** *i)* $l = l_K := |v|^p/(p-1)$ *where* $p > 1$. *Here* $\underline{E} = 0$ *and* $\mu_M$ *is the volume induced by the metric* $g$.

*ii)* $l(x,v) = (1/2)|v|^2 - V(x)$ *where* $V \in C^2(M)$ *(mechanical Lagrangian) . Then* $\underline{E} = \max_{x \in M} V(x)$ *and* $\mu_M$ *of (2.4) is supported at the points of maxima of* $V$.

*iii)* $l(x,v) = l_K(v - \boldsymbol{W}(x))$ *where* $\boldsymbol{W}$ *is a section in* $TM$.
*Then (2.2) implies* $\underline{E} \leq 0$. *In fact, it can be shown that* $\underline{E} = 0$ *for any choice of* $\boldsymbol{W}$.

*iv) In general, if $\boldsymbol{P}$ is in the first cohomology of $M$ ($\mathbf{H}^1(M)$) then $l \mapsto l(x,v) - \langle \boldsymbol{P}, v\rangle$ induced the hamiltonian $h \mapsto h(x, \xi + \boldsymbol{P})$ and $\underline{E} = \alpha(\boldsymbol{P})$ corresponds to the celebrated Mather ($\alpha$) function [20] on the cohomology $\mathbf{H}^1(M)$. See also [27].*

The Monge problem of mass transportation, on the other hand, has a much longer history. Some years before the the French revolution, Monge (1781) proposed to consider the minimal cost of transporting a given mass distribution to another, where the cost of transporting a unit of mass from point $x$ to $y$ is prescribed by a function $C(x,y)$. In modern language, the Monge problem on a manifold $M$ is described as follows: Given a pair of Borel probability measures $\mu_0, \mu_1$ on $M$, consider the set $\mathcal{K}(\mu_0,\mu_1)$ of all Borel mappings $\Phi : M \to M$ transporting $\mu_0$ to $\mu_1$, i.e

$$\Phi \in \mathcal{K}(\mu_0,\mu_1) \Longleftrightarrow \Phi_{\#}\mu_0 = \mu_1$$

and look for the one which minimize the *transportation cost*

$$\mathcal{C}(\mu_0,\mu_1) := \inf_{\Phi}\left\{ \int_M C(x,\Phi(x))d\mu_0(x) \quad ; \quad \Phi \in \mathcal{K}(\mu_0,\mu_1) \right\} . \tag{2.5}$$

In this generality, the set $\mathcal{K}(\mu_0,\mu_1)$ can be empty if, e.g., $\mu_0$ contains an atomic measure while $\mu_1$ does not, so $C(\mu_0,\mu_1) = \infty$ in that case. In 1942, Kantorovich proposed a relaxation of this deterministic definition of the Monge cost. Instead of the (very nonlinear) set $\mathcal{K}(\mu_0,\mu_1)$, he suggested to consider the set $\mathcal{P}(\mu_0,\mu_1)$ defined in section 1.2-(6). Then, the definition of the Monge metric is relaxed into the linear optimization

$$\mathcal{C}(\mu_0,\mu_1) = \min_{\Lambda}\left\{ \int_{M\times M} C(x,y)d\Lambda(x,y) \quad ; \quad \Lambda \in \mathcal{P}(\mu_0,\mu_1) \right\} . \tag{2.6}$$

**Example 2.2.** *The Wasserstein distance $W_p$ ($p \geq 1$) is obtained by the power $p$ of the metric $D$ induced by the Riemannian structure:*

$$W_p(\mu_0,\mu_1) = \min_{\Lambda}\left\{ \left[ \int_{M\times M} D^p(x,y)d\Lambda(x,y) \right]^{1/p} \quad ; \ \Lambda \in \mathcal{P}(\mu_0,\mu_1) \right\} \tag{2.7}$$

The advantage of this relaxed definition is that $C(\mu_0,\mu_1)$ is always finite, and that a minimizer of (2.6) always exists by the compactness of the set $\mathcal{P}(\mu_0,\mu_1)$ in the weak topology $C^*(M \times M)$. If $\mu_0$ contains no atomic points then it can be shown that $C(\mu_0,\mu_1)'s$ given by (2.5) and (2.6) coincide [1].

The theory of Monge-Kantorovich (M-K) was developed in the last few decades in a countless number of publications. For updated reference see [12], [28]. [2]

Returning now to WKAM, it was observed by Bernard and Buffoni ([4][5]- see also [29]) that the minimal measure and the ground energy can be expressed in terms of the M-K problem subjected to the cost function induced by the Lagrangian (recall section 1.2-5)

$$C_T(x,y) := \inf_{\boldsymbol{z}}\left\{ \int_0^T l\left(\boldsymbol{z}(s);\dot{\boldsymbol{z}}(s)\right)ds \quad , \ \boldsymbol{z} \in \mathcal{K}^T_{x,y} \right\} , T > 0 . \tag{2.8}$$

[2]By convention, the name "Monge problem" is reserved for the metric cost, while "Monge-Kantorovich problem" is usually referred to general cost functions

Then

$$\mathcal{C}_T(\mu) := \mathcal{C}_T(\mu, \mu) = \min_{\Lambda} \left\{ \int_{M \times M} C_T(x,y) d\Lambda(x,y) \quad ; \quad \Lambda \in \mathcal{P}(\mu, \mu) \right\}$$

and

$$\min_{\mu} \left\{ \mathcal{C}_T(\mu) \ ; \ \mu \in \mathcal{M}_1^+ \right\} = -T\underline{E} \tag{2.9}$$

where the minimizers of (2.9) coincide, for any $T > 0$, with the projected Mather measure $\mu_M$ maximizing (2.4) [5]. The action $C_T$ induces a metric on the manifold $M$:

$$(x,y) \in M \times M \mapsto D_E(x,y) = \inf_{T>0} C_T(x,y) + TE \ . \tag{2.10}$$

**Example 2.3.**

*i) For* $l(x,v)=|v|^p/(p-1)$, $p>1$ *we get* $C_T(x,y)=D(x,y)^p/(p-1)T^{p-1}$ *while* $D_E(x,y)=pE^{1-1/p}D_g(x,y)/(p-1)$ *if* $E\geq 0$, $D_E(x,y)=-\infty$ *if* $E<0$.

*ii)* $l(x,v)=(1/2)|v|^2-V(x)$ *where* $V\in C^2(M)$ *(mechanical Lagrangian) . Then* $D_E(x,y)$ *is the geodesic distance induced by conformal equivalent metric* $(M,(E-V)g)$ *on* $M$*, where* $E\geq \underline{E}=\sup_M V$*.*

It is not difficult to see that either $D_E(x,x)=0$ for any $x\in M$, or $D_E(x,y)=-\infty$ for any $x,y\in M$. In fact, it follows ([22], [10]) that $D_E(x,y)=-\infty$ for $E<\underline{E}$ and $D_E(x,x)=0$ for $E\geq\underline{E}$ and any $x,y\in M$.

Let now $\lambda^+$ , $\lambda^-\in\mathcal{M}^+$ where $\lambda:=\lambda^+-\lambda^-\in\mathcal{M}_0$, that is $\int_M d\lambda=0$. Let

$$\mathcal{D}_E(\lambda):=\mathcal{D}_E(\lambda^+,\lambda^-)=\min_{\Lambda}\left\{\int_{M\times M}D_E(x,y)d\Lambda(x,y)\quad;\ \Lambda\in\mathcal{P}(\lambda)\right\} \tag{2.11}$$

be the Monge distance of $\lambda^+$ and $\lambda^-$ with respect to the metric $D_E$. There is a dual formulation of $\mathcal{D}_E$ as follows: Consider the set $\mathcal{L}_E$ of $D_E$ Lipschitz functions on $M$:

$$\mathcal{L}_E:=\{\phi\in C(M)\ ;\quad \phi(x)-\phi(y)\leq D_E(x,y)\quad\forall\ x,y\in M\} \tag{2.12}$$

Then (see, e.g [12], [26])

$$\mathcal{D}_E(\lambda)=\max_{\phi}\left\{\int_M\phi d\lambda\quad;\ \phi\in\mathcal{L}_E.\right\} \tag{2.13}$$

# 3 Description of the main results

The object of this paper is to establish some relations between the action $C_T$ and a modified action $\widehat{\mathcal{C}}_T$ defined below.

## 3.1 Unconditional action

For given $\lambda\in\mathcal{M}_0$ we generalize (2.1) into

$$\mathcal{M}_\lambda:=\left\{\nu\in\mathcal{M}_1^+(TM)\ ;\quad\int_{TM}l(x,v)d\nu(x,v)<\infty\quad;\quad\int_{TM}\langle d\phi,v\rangle d\nu=\int_M\phi d\lambda\ \text{ for any } \phi\in C^1(M)\right\} \tag{3.1}$$

and define

$$\widehat{\mathcal{C}}(\lambda):=\inf_{\nu}\left\{\int_{TM}l(x,v)d\nu(x,v)\quad;\quad\nu\in\mathcal{M}_\lambda\right\}\ . \tag{3.2}$$

The modified action $\widehat{\mathcal{C}}_T:\mathcal{M}_0\to\mathbb{R}\cup\{\infty\}$, $T>0$ have several equivalent definitions as given in Theorem 1 below:

**Theorem 1.** *The following definitions are equivalent:*

1. $\widehat{\mathcal{C}}_T(\lambda):=T\widehat{\mathcal{C}}\left(\frac{\lambda}{T}\right)$ .

2. $\widehat{\mathcal{C}}_T(\lambda) := \min_\mu \sup_\phi \left\{ \int_M -Th(x, d\phi) d\mu + \phi d\lambda \;\; ; \;\; \mu \in \mathcal{M}_1^+, \;\; \phi \in C^1(M) \right\}$ .

3. $\widehat{\mathcal{C}}_T(\lambda) := \max_{E \geq \underline{E}} \left[ \mathcal{D}_E(\lambda) - ET \right]$ .

*In addition if* $T_c := D_{\underline{E}}^{'+}(\lambda) < \infty$ *then for* $T \geq T_c$*,*

$$\widehat{\mathcal{C}}_T(\lambda) = \widehat{\mathcal{C}}_{T_c}(\lambda) - T\underline{E} \ .$$

*In that case the minimizer* $\mu_\lambda^T \in \mathcal{M}_1^+$ *of (3),* $T > T_c$ *is given by*

$$\mu_\lambda^T = \frac{T_c}{T} \mu_\lambda^{T_c} + \left( 1 - \frac{T_c}{T} \right) \mu_M \ ,$$

*where* $\mu_M$ *is the projected Mather measure.*

**Remark 3.1.** *Note that* $\mathcal{D}_E(\lambda)$ *(2.11, 2.13) is a monotone non-decreasing and concave function of* $E$ *while* $\mathcal{D}_{\underline{E}}(\lambda) > -\infty$ *by definition. Hence the right-derivative of* $\mathcal{D}_E^{'+}(\lambda)$ *as a function of* $E$ *is defined and positive (possibly* $+\infty$ *at* $E = \underline{E}$*).*

**Remark 3.2.** *A special case of Theorem 1 was introduced in [30].*

For the next result we need a two technical assumptions:

$\mathbf{H_1}$ There exists a sequence of smooth, positive mollifiers $\delta_\varepsilon : M \times M \to \mathbb{R}^+$ such that, for any $\phi \in C^0(M)$ (res. $\phi \in C^1(M)$)

$$\lim_{\varepsilon \to 0} \delta_\varepsilon * \phi = \phi$$

where the convergence is in $C^0(M)$ (res. $C^1(M)$) and for any $\varepsilon > 0$ and $\phi \in C^1(M)$

$$\delta_\varepsilon * d\phi = d(\delta_\varepsilon * \phi) \ .$$

$\mathbf{H_2}$ For any $(x, p) \in T^*M$ and $\varepsilon > 0$ there exists $\delta > 0$ such that $h(x, \xi) - h(y, \xi_y) \leq \varepsilon(h(x, \xi) + 1)$ provided $D(x, y) < \delta$. Here $\xi_y$ is obtained by parallel translation of $(x, \xi)$ to $y$.

**Remark 3.3.** $\mathbf{H_1}$ *holds for homogeneous spaces, e.g the flat* $d-$*torus* $\mathbb{R}^d / \mathbb{Z}^n$ *or the sphere* $\mathbb{S}^{d-1} = SO(d)/SO(1)$*.*
$\mathbf{H_2}$ *holds, in particular, for any mechanical hamiltonian with continuous potential.*

**Theorem 2.** *Assume* $\mathbf{H_1} + \mathbf{H_2}$*. For any* $\lambda = \lambda^+ - \lambda^-$ *where* $\lambda^\pm \in \mathcal{M}_1^+$*,*

$$\widehat{\mathcal{C}}_T(\lambda) = \lim_{\varepsilon \to 0} \min_{\mu \in \mathcal{M}_1^+} \varepsilon^{-1} \mathcal{C}_{\varepsilon T}(\mu + \varepsilon \lambda^-, \mu + \varepsilon \lambda^+) \ .$$

As an application of Theorem 2 we may consider the case where the lagrangian $l$ is homogeneous with respect to a Riemannian metric $g_{(x)}$:

**Example 3.1.** *If* $l(x,v) = |v|^p/(p-1)$ *where* $p > 1$*. Then* $C_T(x,y) = \frac{D^p(x,y)}{(p-1)T^{p-1}}$ *while* $D_E(x,y) = \frac{p}{p-1}E^{(p-1)/p}D(x,y)$ *and* $\underline{E} = 0$*. It follows that*

$$\widehat{\mathcal{C}}_T(\lambda) = \frac{W_1^p(\lambda)}{(p-1)T^{p-1}}, \quad \varepsilon^{-1}\mathcal{C}_{\varepsilon T}(\mu + \varepsilon\lambda^+, \mu + \varepsilon\lambda^-) = \frac{W_p^p(\mu + \varepsilon\lambda^+, \mu + \varepsilon\lambda^-)}{(p-1)T^{p-1}\varepsilon^p} \tag{3.3}$$

*where the Wasserstein distance* $W_p$ *is defined in (2.7). Hence, by Theorem 1 and Theorem 2*

$$W_1(\lambda) = \lim_{\varepsilon \to 0} \varepsilon^{-1} \inf_{\mu \in \mathcal{M}_1^+} W_p(\mu + \varepsilon\lambda^-, \mu + \varepsilon\lambda^+) \ .$$

**Remark 3.4.** *The optimal transport description of the weak KAM theory (2.9) can be considered as a special case of Theorem 2 where* $\lambda = 0$*. Indeed* $\inf_{\mu \in \mathcal{M}_1^+} \varepsilon^{-1} C_{\varepsilon T}(\mu, \mu) = -T\underline{E}$ *by (2.9). On the other hand, since* $\mathcal{D}_E(0) = 0$ *for any* $E \geq \underline{E}$ *it follows that* $T_c = 0$*, hence* $\widehat{\mathcal{C}}_{T_c}(0) = 0$ *so* $\widehat{\mathcal{C}}_T(0) = -T\underline{E}$ *as well by the last part of Theorem 1.*

### 3.2 Conditional action

There is also an interest in the definition of action (and metric distance) conditioned with a given probability measure $\mu \in \mathcal{M}_1^+$. We introduce these definitions and reformulate parts of the main results Theorems 1-2 in terms of these.

For a given $\mu \in \mathcal{M}_1^+$ and $E \geq \underline{E}$, let

$$\mathcal{H}_E(\mu) := \left\{ \phi \in C^1(M) \ ; \ \int_M h(x, d\phi) d\mu \leq E \right\} \ . \tag{3.4}$$

In analogy with (2.13) we define the $\mu-$*conditional metric* on $\lambda \in \mathcal{M}_0$:

$$\mathcal{D}_E(\lambda \| \mu) := \sup_\phi \left\{ \int_M \phi d\lambda \ \ ; \ \ \phi \in \mathcal{H}_E(\mu) \right\} \ . \tag{3.5}$$

The *conditioned, modified action* with respect to $\mu \in \mathcal{M}_1^+$ is defined in analogy with Theorem 1 (2, 3)

$$\widehat{\mathcal{C}}_T(\lambda \| \mu) := \max_{E \geq \underline{E}} \mathcal{D}_E(\lambda \| \mu) - ET \equiv \sup_{\phi \in C^1(M)} \int_M -Th(x, d\phi) d\mu + \phi d\lambda \ . \tag{3.6}$$

**Example 3.2.** *As in Example 3.1,* $l(x,v) = |v|^p/(p-1)$ *implies* $h(\xi) = q^{-q}|\xi|^q$ *where* $q = p/(p-1)$*. Then (3.4, 3.5) is related to (1.4), that is* $W_1^{(p)}(\lambda \| \mu) = \mathcal{D}_E(\lambda \| \mu)$ *where* $E = q^{-q}$ *or*

$$\mathcal{D}_E(\lambda \| \mu) = qE^{1/q} W_1^{(p)}(\lambda \| \mu) \ , \ \ \widehat{\mathcal{C}}_T(\lambda \| \mu) = \frac{q-1}{T^{1/(q-1)}} \left( W_1^{(p)}(\lambda \| \mu) \right)^p \tag{3.7}$$

**Remark 3.5.** *It seems there is a relation between this definition and the tangential gradient [6]. There are also possible applications to optimal network and irrigation theory, where one wishes to minimize* $D(\lambda \| \mu)$ *over some constrained set of* $\mu \in \mathcal{M}_1^+$ *(the irrigation network) for a prescribed* $\lambda$ *(representing the set of sources and targets). See, e.g. [8], [9] and the ref. within.*

The next result is

**Theorem 3.** *For any* $\lambda \in \mathcal{M}_0$,

$$\mathcal{D}_E(\lambda) = \min_{\mu \in \mathcal{M}_1^+} \mathcal{D}_E(\lambda \| \mu) \ , \quad \widehat{\mathcal{C}}_T(\lambda) = \min_{\mu \in \mathcal{M}_1^+} \widehat{\mathcal{C}}_T(\lambda \| \mu) \ .$$

The analog of Theorem 2 holds for the conditional action as well. However, we can only prove the $\Gamma-$convergence in that case. Recall that a sequence of functionals $F_n : \mathbf{X}_n \to \mathbb{R} \cup \{\infty\}$ is said to $\Gamma-$converge to $F : \mathbf{X} \to \mathbb{R} \cup \{\infty\}$ ($\Gamma - \lim_{n\to\infty} F_n = F$) if and only if

(i) $\mathbf{X}_n \subset \mathbf{X}$ for any $n$.

(ii) For any sequence $x_n \in \mathbf{X}_n$ converging to $x \in \mathbf{X}$ in the topology of $\mathbf{X}$ ,

$$\liminf_{n\to\infty} F_n(x_n) \geq F(x) \ .$$

(iii) For any $x \in \mathbf{X}$ there exists a sequence $\hat{x}_n \in \mathbf{X}_n$ converging to $x \in \mathbf{X}$ in the topology of $\mathbf{X}$ for which

$$\lim_{n\to\infty} F_n(\hat{x}_n) = F(x) \ .$$

In Theorem 4 below the $\Gamma-$convergence is related to the special case where $\mathbf{X}_n = \mathbf{X}$:

**Theorem 4.** *Let* $\mathbf{X}_n = \mathcal{M}_0 \times \mathcal{M}_1^+ = \mathbf{X}$ *and* $F_n(\lambda, \mu) := n\mathcal{C}_{T/n}(\mu + \lambda^-/n, \mu + \lambda^+/n)$. *Then*

$$\widehat{\mathcal{C}}_T(\cdot \| \cdot) = \Gamma - \lim_{n\to\infty} F_n \ .$$

From Theorem 4 and Theorem 2 it follows immediately

**Corollary 3.1.** *In addition, if* $\mu_n$ *is a minimizer of* $F_n$ *in* $\mathcal{M}_1^+$ *then any converging subsequence of* $\mu_n$, $n \to \infty$, *converges to a minimizer of* $\widehat{\mathcal{C}}(\lambda \| \cdot)$ *in* $\mathcal{M}_1^+$.

Finally, we note that (1.7) is a special case of Theorem 4. Using Examples 3.1, 3.2 with $\varepsilon = 1/n$, recalling $(q-1)^{-1} = p - 1$ we obtain

**Corollary 3.2.**

$$W_1(\lambda) = \lim_{n\to\infty} n \min_{\mu \in \mathcal{M}_1^+} W_p(\mu + \lambda^+/n, \mu + \lambda^-/n)$$

# 4 Proof of Theorems 1&3

We first show that $\widehat{\mathcal{C}}(\lambda) < \infty$ (recall (3.2)).

**Lemma 4.1.** *For any* $\lambda \in \mathcal{M}_0$, $\mathcal{M}_\lambda \neq \emptyset$. *In particular, since the Lagrangian* $l$ *is bounded from below,* $\widehat{\mathcal{C}}(\lambda) < \infty$.

*Proof.* It is enough to show that there exists a compact set $K \subset TM$ and a sequence $\{\lambda_n\} \subset \mathcal{M}_0$ converging weakly to $\lambda$ such that for each $n$ there exists $\nu_n \in \mathcal{M}_{\lambda_n}$ whose support is contained in $K$. Indeed, such a set is compact and there exists a weak limit $\nu = \lim_{n\to\infty} \nu_n$ which satisfies $\lim_{n\to\infty} v\nu_n = v\nu$ as well. Hence, if $\phi \in C^1(M)$ then

$$\lim_{n\to\infty} \int_M \langle d\phi, v\rangle d\nu_n = \int_M \langle d\phi, v\rangle d\nu \quad , \quad \lim_{n\to\infty} \int_M \phi d\lambda_n = \int_M \phi d\lambda_n \ .$$

Since $\nu_n \in \mathcal{M}_{\lambda_n}$ we get

$$\int_M \langle d\phi, v\rangle d\nu_n = \int_M \phi d\lambda_n$$

for any $n$, so the same equality holds for $\nu$ as well.

Now, we consider

$$\lambda_n = \alpha_n \sum_{j=1}^{n} \left(\delta_{x_j} - \delta_{y_j}\right) \tag{4.1}$$

where $x_j, y_j \in M$ and $\alpha_n > 0$. For any pair $(x_j, y_j)$ consider a geodesic arc corresponding to the Riemannian metric which connect $x$ to $y$, parameterized by the arc length: $\boldsymbol{z}_j : [0,1] \to M$ and $|\dot{\boldsymbol{z}}| = D(x_j, y_j)$ (recall section 1.2-(1)). Then

$$\nu_n := \alpha_n \sum_{j=1}^{n} \int_0^1 \delta_{x-\boldsymbol{z}_j(t), v-\dot{\boldsymbol{z}}_j(t)} dt$$

satisfies for any $\phi \in C^1(M)$

$$\int_M \langle d\phi, v\rangle d\nu_n = \alpha_n \sum_{j=1}^{n} \int_0^1 \langle d\phi\left(\boldsymbol{z}_j(s), \dot{\boldsymbol{z}}_j(s)\right) \dot{\boldsymbol{z}}_j(t)\rangle dt = \alpha_n \sum_{j=1}^{n} \int_0^1 \frac{d}{dt}\phi\left(\boldsymbol{z}_j(s)\right) dt$$
$$= \alpha_n \sum_{j=1}^{n} [\phi(y_j) - \phi(x_j)] = \int_M \phi d\lambda_n \tag{4.2}$$

hence $\nu_n \in \mathcal{M}_{\lambda_n}$. Finally, we can certainly find such a sequence $\lambda_n$ of the form (4.1) which converges weakly to $\lambda$. □

## 4.1 Point distances and Hamiltonians

For $E \in \mathbb{R}$, let $\sigma_E : TM \to \mathbb{R}$ the support function of the level surface $h(x,\xi) \leq E$, that is:

$$\sigma_E(x,v) := \sup_{\xi \in T_x^* M} \left\{ \langle \xi, v\rangle_{(x)} \ ; \ h(x,\xi) \leq E \right\} \ . \tag{4.3}$$

It follows from our standing assumptions (Section 1.2-7) that $\sigma_E$ is differentiable as a function of $E$ for any $(x,v) \in TM$. For the following Lemma see e.g. [25].

Recall that

$$D_E(x,y) := \inf_{T>0} C_T(x,y) + ET \tag{4.4}$$

where $C_T$ as defined in (2.8). Recall also section 1.2-5:

**Lemma 4.2.** .

$$D_E(x,y) = \inf_{\boldsymbol{z}\in\mathcal{K}^1_{x,y}} \int_0^1 \sigma_E\left(\boldsymbol{z}(s), \dot{\boldsymbol{z}}(s)\right) ds \ . \tag{4.5}$$

Given $x \in M$, let

$$\underline{E} := \inf\left\{E \in \mathbb{R};\ D_E(x,x) > -\infty\right\} \tag{4.6}$$

For the following Lemma see [21] (also [27]):

**Lemma 4.3.** *$\underline{E}$ is independent of $x \in M$. The definitions (4.6) and (2.2) and (2.4) are equivalent. If $E \geq \underline{E}$ then $D_E(x,y) > -\infty$ for any $x, y \in M$ and, in addition*

*i) $D_E(x,x) = 0$ for any $x \in M$.*

*ii) For any $x, y, z \in M$, $D_E(x,z) \leq D_E(x,y) + D_E(y,z)$*

From (4.4), Lemma 4.2 and the continuity of $\sigma_E$ with respect to $E \geq \underline{E}$ we get

**Corollary 4.1.** *If $E \geq \underline{E}$ then for any $x, y \in M$, $D_E(x,y)$ is continuous, monotone non-decreasing and concave as a function of $E$.*

Note that the differentiability of $\sigma_E$ with respect to $E$ does *not* imply that $D_E(x,y)$ is differentiable for each $x, y \in M$. However, since $D_E(x,y)$ is a concave function of $E$ for each $x, y \in M$, it is differentiable for Lebesgue almost any $E > \underline{E}$. We then obtain by differentiation

**Lemma 4.4.** *If $E$ is a point of differentiability of $D_E(x,y)$ then there exists a geodesic arc $\boldsymbol{z} \in \mathcal{K}^1_{x,y}$ realizing (4.5) such that the $E$ derivative of $D_E(x,y)$ is given by*

$$T_E(x,y) := \frac{d}{dE} D_E(x,y) = \int_0^1 \sigma^{'}_E\left(\boldsymbol{z}(s), \dot{\boldsymbol{z}}(s)\right) ds \ , \tag{4.7}$$

*where $\sigma^{'}_E$ is the $E$ derivative of $\sigma_E$. Moreover*

$$D_E(x,y) = C_{T_E(x,y)}(x,y) + E T_E(x,y) \ . \tag{4.8}$$

From (4.3) we get $\sigma_E(x,v) \leq |v| \max\{|p| \ ; \ h(x,\xi) \leq E\}$. From our standing assumption on $h$ (section 1.2-(7)) and (4.5) we obtain

**Lemma 4.5.** *For any $x, y \in M$ and $E \geq \underline{E}$*

$$D_E(x,y) \leq \hat{h}^{-1}(E + C) D(x,y)$$

*In particular*

$$\lim_{E\to\infty} E^{-1} D_E(x,y) = 0 \tag{4.9}$$

*uniformly on $M \times M$.*

**Corollary 4.2.** *For $E \geq \underline{E}$, the set $\mathcal{L}_E$ (2.12) is contained in the set of Lipschitz functions with respect to $D$, and $\mathcal{L}_E$ is locally compact in $C(M)$.*

Given $\phi \in C^1(M)$ let

$$\overline{H}(\phi) := \sup_{x\in M} h(x, d\phi) \quad . \tag{4.10}$$

We extend the definition of $\overline{H}$ to the larger class of Lipschitz functions by the following

**Lemma 4.6.** *If $\phi \in C^1(M)$ then*

$$\overline{H}(\phi) = \min_{E\geq \underline{E}} \{E;\ \phi \in \mathcal{L}_E\} \ ,$$

*where $\mathcal{L}_E$ as defined in (2.12).*

*Proof.* First we show that if $\phi \in \mathcal{L}_E \cap C^1(M)$ then $h(x, d\phi) \leq E$ for all $x \in M$. Indeed, for any $x, y \in M$ and any curve $z(\cdot)$ connecting $x$ to $y$

$$\phi(y) - \phi(x) = \int_0^1 d\phi(z(t)) \cdot \dot{z} dt \leq D_E(x, y) \leq \int_0^1 \sigma_E(z(t), \dot{z}(t))dt$$

hence $d\phi(x) \cdot v \leq \sigma_E(x, v)$ for any $v \in T_xM$. Then, by definition, $d\phi(x)$ is contained in any supporting half space which contains the set $Q_x(E) := \{\xi \in T^*_xM;\ h(x, \xi) \leq E\}$. Since this set is convex by assumption, it follows that $d\phi \in Q_x(E)$, so $h(x, d\phi) \leq E$ for any $x \in M$. Hence $\overline{H}(\phi) \leq E$.

Next we show the opposite inequality $h(x, d\phi) \geq E$ for all $x \in M$. Recall (4.8). Then for any $\varepsilon > 0$ we can find $T_\varepsilon > 0$ and $\boldsymbol{z}_\varepsilon \in \mathcal{K}^{T_\varepsilon}_{x,y}$ so

$$D_E(x, y) \geq \int_0^{T_\varepsilon} l(\boldsymbol{z}_\varepsilon(t), \dot{\boldsymbol{z}}_\varepsilon(t))dt + (E - \varepsilon)T_\varepsilon \ . \tag{4.11}$$

Next, for a.e $t \in [0, T_\varepsilon]$

$$h\left(\boldsymbol{z}_\varepsilon(t), d\phi(\boldsymbol{z}_\varepsilon(t))\right) \geq \dot{\boldsymbol{z}}_\varepsilon(t) \cdot d\phi(\boldsymbol{z}_\varepsilon(t)) - l\left(\boldsymbol{z}_\varepsilon(t), \cdot \boldsymbol{z}_\varepsilon(t)\right) \ . \tag{4.12}$$

Integrate (4.12) from 0 to $T_\varepsilon$ and use $\boldsymbol{z}_\varepsilon \in \mathcal{K}^{T_\varepsilon}_{x,y}$, (4.11, 4.12) and the definition of $\mathcal{L}_E$ to obtain

$$T_\varepsilon^{-1} \int_0^{T_\varepsilon} h\left(\boldsymbol{z}_\varepsilon(t), d\phi(\boldsymbol{z}_\varepsilon(t))\right) dt \geq T_\varepsilon^{-1} \left[\phi(y) - \phi(x)\right] - T_\varepsilon^{-1} \int_0^{T_\varepsilon} l\left(\boldsymbol{z}_\varepsilon(t), \cdot \boldsymbol{z}_\varepsilon(t)\right) dt \geq E - \varepsilon \ .$$

Hence, the supremum of $h(x, d\phi)$ along the orbit of $\boldsymbol{z}_\varepsilon$ is, at least, $E - \varepsilon$. Since $\varepsilon$ is arbitrary, then $\overline{H}(\phi) \geq E$. □

## 4.2 Measure distances and Hamiltonians

From Lemma 4.6 and Corollary 4.2 we extend the definition of $\overline{H}$ to the space $Lip(M)$ of Lipschitz functions on $M$. Let now define for $\lambda \in \mathcal{M}_0$

$$\overline{H}^*_T(\lambda) := \sup_{\phi\in Lip(M)} \left\{-T\overline{H}(\phi) + \int_M \phi d\lambda\right\} \in \mathbb{R} \cup \{\infty\} \ . \tag{4.13}$$

**Proposition 4.1.** *For any $\lambda \in \mathcal{M}_0$*

$$\overline{H}^*_T(\lambda) = \sup_{E \geq \underline{E}} \{\mathcal{D}_E(\lambda) - TE\} \ . \tag{4.14}$$

*Proof.* By definition of $\overline{H}^*$ and Lemma 4.6,

$$\begin{aligned}\overline{H}^*_T(\lambda) = \sup_{\phi \in Lip(M)} \left[ \int_M \phi d\lambda - T\overline{H}(\phi) \right] = \sup_{\phi \in Lip(M)} \sup_{E \geq \underline{E}} \left[ \int_M \phi d\lambda - TE \ ; \ \phi \in \mathcal{L}_E \right] \\ = \sup_{E \geq \underline{E}} \sup_{\phi \in Lip(M)} \left[ \int_M \phi d\lambda - TE \ ; \ \phi \in \mathcal{L}_E \right] = \sup_{E \geq \underline{E}} \{\mathcal{D}_E(\lambda) - TE\} , \end{aligned} \tag{4.15}$$

where we used the duality relation given by (2.13). □

**Corollary 4.3.** *$\overline{H}^*_T$ is weakly continuous on $\mathcal{M}_0$.*

*Proof.* For each $E \geq \underline{E}$, the Monge-Kantorovich metric $\mathcal{D}_E : \mathcal{M}_0 \to \mathbb{R}$ is continuous on $\mathcal{M}_0$ (under weak* topology). Indeed, it is u.s.c. by (2.11) and l.s.c. by the dual formulation (2.13).

Also, for each $\lambda \in \mathcal{M}^+_1$, $\mathcal{D}_E(\lambda)$ is concave and finite in $E$ for $E \geq \underline{E}$. It follows that $\mathcal{D}$ is mutually continuous on $[\underline{E}, \infty[ \times \mathcal{M}_0$. From (4.9) we also get that $\mathcal{D}$ is coercive on $\mathcal{M}_0$, that is $\lim_{E\to\infty} E^{-1}\mathcal{D}_E(\lambda) = 0$ locally uniformly on $\mathcal{M}_0$. These imply that $\overline{H}^*_T$ is continuous on $\mathcal{M}_0$ via (4.14). □

We return now to Corollary 4.1 and Lemma 4.4. It follows that for any countable dense set $A \subset M$ there exists a (possibly empty) set $N \subset ]\underline{E}, \infty[$ of zero Lebesgue measure such that $D_E(x, y)$ is differentiable in $E \in ]\underline{E}, \infty[ - N$, for *any* $x, y \in A$. Let $\mathcal{M}(A) \subset \mathcal{M}_0$ be the set of all measures in $\mathcal{M}_0$ which are supported on a *finite* subset of $A$, and such that $\lambda(\{x\})$ is rational for any $x \in A$. Again, since $\mathcal{M}(A)$ is countable, it follows by Corollary 4.1 that $\mathcal{D}_E(\lambda)$ is differentiable (as a function of $E$) for any $\lambda \in \mathcal{M}(A)$ and any $E \in ]\underline{E}, \infty[ - N$ for a (perhaps larger) set $N$ of zero Lebesgue measure. It is also evident that $\mathcal{M}_0$ is the weak closure of $\mathcal{M}(A)$.

**Lemma 4.7.** *For any $\lambda^+ - \lambda^- \equiv \lambda \in \mathcal{M}(A)$ and $E \in ]\underline{E}, \infty[ - N$, there exists an optimal plan $\Lambda^\lambda_E \in \mathcal{P}(\lambda^+, \lambda^-)$ realizing*

$$\int_{M\times M} D_E(x, y) d\Lambda^\lambda_E(x, y) = \min_{\Lambda \in \mathcal{P}(\lambda^+,\lambda^-)} \int_{M\times M} D_E(x, y) d\Lambda(x, y) \equiv \mathcal{D}_E(\lambda) \tag{4.16}$$

*for which*

$$\frac{d}{dE}\mathcal{D}_E(\lambda) = \sum_{x,y\in A} \Lambda^\lambda_E(\{x, y\}) T_E(x, y) \ . \tag{4.17}$$

*Proof.* Let $E_n \searrow E$. For each $n$, set $\Lambda^\lambda_{E_n}$ be a minimizer of (4.16) subjected to $E = E_n$. We choose a subsequence so that the limit

$$\Lambda^\lambda_{E^+}(\{x, y\}) := \lim_{n\to\infty} \Lambda^\lambda_{E_n}(\{x, y\}) \tag{4.18}$$

exists for any $x, y \in A$. Evidently, $\Lambda^{\lambda}_{E+} \in \mathcal{P}(\lambda^+, \lambda^-)$ is an optimal plan for (4.16). Next,

$$\mathcal{D}_{E_n}(\lambda) - \mathcal{D}_E(\lambda) \geq \sum_{x,y\in A} \Lambda^{\lambda}_{E_n}(\{x,y\}) \left(D_{E_n}(x,y) - D_E(x,y)\right)$$

Divide by $E_n - E > 0$ and let $n \to \infty$, using (4.18) and (4.7) we get

$$\frac{d}{dE}\mathcal{D}_E(\lambda) \geq \sum_{x,y\in A} \Lambda^{\lambda}_{E+}(\{x,y\}) T_E(x,y) \ . \tag{4.19}$$

We repeat the same argument for a sequence $E^n \nearrow E$ for which

$$\Lambda^{\lambda}_{E^-}(\{x,y\}) := \lim_{n\to\infty} \Lambda^{\lambda}_{E_n}(\{x,y\})$$

and get

$$\frac{d}{dE}\mathcal{D}_E(\lambda) \leq \sum_{x,y\in A} \Lambda^{\lambda}_{E^-}(\{x,y\}) T_E(x,y) \ . \tag{4.20}$$

Again $\Lambda^{\lambda}_{E^-}$ is an optimal plan as well. If $\Lambda^{\lambda}_{E^-} = \Lambda^{\lambda}_{E^+}$ then we are done. Otherwise, define $\Lambda^{\lambda}_{E^-}$ as a convex combination of $\Lambda^{\lambda}_{E^-}$ and $\Lambda^{\lambda}_{E^+}$ for which the equality (4.17) holds due to (4.19, 4.20). □

Given $x, y \in M$, let $E$ be a point of differentiability of $D_E(x,y)$, and $\boldsymbol{z}^E_{x,y} : [0,1] \to M$ a geodesic arc connecting $x, y$ and realizing (4.7). Then $d\tau^E_{x,y} := \sigma'_E\left(\boldsymbol{z}^E_{x,y}, \dot{\boldsymbol{z}}^E_{x,y}\right) ds$ is a non-negative measure on $[0,1]$, and (4.7) reads $T_E(x,y) = \int_0^1 d\tau^E_{x,y}$. Let $\mu^E_{x,y}$ be the measure on $M$ obtained by pushing $\tau^E_{x,y}$ from $[0,1]$ to $M$ via $\boldsymbol{z}^E_{x,y}$:

$$\mu^E_{x,y} := \left(\boldsymbol{z}^E_{x,y}\right)_{\#} \tau^E_{x,y} \in \mathcal{M}^+ \ ,$$

that is, for any $\phi \in C(M)$,

$$\int_M \phi d\mu^E_{x,y} := \int_0^1 \phi\left(\boldsymbol{z}^E_{x,y}(t)\right) d\tau^E_{x,y} \ . \tag{4.21}$$

**Definition 4.1.** *For any $\lambda \in \mathcal{M}(A)$ and $E \in ]\underline{E}, \infty[-N$ let*

$$\mu^E_{\Lambda} := \sum_{x,y\in A} \Lambda^{\lambda}_E(\{x,y\}) \mu^E_{x,y}$$

*where $\mu^E_{x,y}$ are as given in (4.21) and $\Lambda^{\lambda}_E$ is the particular optimal plan given in Lemma 4.7.*

**Remark 4.1.** *Note that $\int_M d\mu^E_{\Lambda} = \mathcal{D}'_E(\lambda)$ for any $\lambda \in \mathcal{M}_0(A)$ and $E \in ]\underline{E}, \infty[-N$ by Lemma 4.7, where $\mathcal{D}'_E(\lambda) = (d/dE)\mathcal{D}_E(\lambda)$.*

**Definition 4.2.** *For any $\lambda \in \mathcal{M}_0$, $T > 0$, $E(\lambda, T)$ is the maximizer of (4.14), that is*

$$\mathcal{D}_{E(\lambda,T)}(\lambda) - TE(\lambda,T) \equiv \overline{H}^*_T(\lambda) \ .$$

By Corollary 4.1 (in particular, the concavity of $\mathcal{D}_E(\lambda)$ with $E$) we obtain

**Lemma 4.8.** *If $E(\lambda,T) > \underline{E}$ then*

$$\frac{d^+}{dE}\mathcal{D}_E(\lambda)\Big|_{E=E(\lambda,T)} \leq T \leq \frac{d^-}{dE}\mathcal{D}_E(\lambda,T)\Big|_{E=E(\lambda,T)}$$

*where $d^+/dE$ (res. $d^-/dE$) stands for the right (res. left) derivative. If $E(\lambda,T) = \underline{E}$ then*

$$\frac{d^+}{dE}\mathcal{D}_E(\lambda)\Big|_{E=\underline{E}} \leq T \ .$$

### 4.3 Proof of Theorem 1 (1 $\leftrightarrows$2)

First we note that it is enough to assume $T = 1$. Consider

$$\mathcal{F}(\mu,\phi) := \int_M -h(x,d\phi)d\mu + \phi d\lambda \tag{4.22}$$

where $\lambda \in \mathcal{M}_0$ is prescribed. Evidently, $\mathcal{F}$ is convex lower semi continuous (l.s.c) in $\mu$ on $\mathcal{M}_1^+$ and concave upper semi continuous (u.s.c) in $\phi$ on $C^1(M)$. Since $\mathcal{M}_1^+$ is compact, the Minimax Theorem implies

$$\sup_{\phi\in C^1(M)} \min_{\mu\in\mathcal{M}_1^+} \mathcal{F}(\mu,\phi) = \min_{\mu\in\mathcal{M}_1^+} \sup_{\phi\in C^1(M)} \mathcal{F}(\mu,\phi) \ . \tag{4.23}$$

Next define

$$\mathcal{G}(\nu,\phi) := \int_{TM} \left(l(x,v) - \langle d\phi, v\rangle\right) d\nu + \int_M \phi d\lambda \ .$$

on $\mathcal{M}_1^+(TM) \times C^1(M)$. Then (recall (3.1))

$$\sup_{\phi\in C^1(M)} \inf_{\nu\in\mathcal{M}_1^+(TM)} \mathcal{G}(\nu,\phi) \leq \inf_{\nu\in\mathcal{M}_\lambda} \int_{TM} l(x,v)d\nu \equiv \widehat{\mathcal{C}}(\lambda) \ . \tag{4.24}$$

Now

$$\overline{\mathcal{G}}(\nu) := \sup_{\phi\in C^1(M)} \mathcal{G}(\nu,\phi) \equiv \begin{cases} \int_{TM} l(x,v)d\nu & \text{if } \nu \in \mathcal{M}_\lambda \\ \infty & \text{if } \nu \notin \mathcal{M}_\lambda \end{cases} \ .$$

We recall, again, from the Minmax Theorem that the inequality in (4.24) turns into an equality provided the set $\{\nu \in \mathcal{M}_1^+(TM);\ \overline{\mathcal{G}}(\nu) \leq \widehat{\mathcal{C}}(\lambda)\}$ is compact. However $\widehat{\mathcal{C}}(\nu) < \infty$ by Lemma 4.1. Since $l$ is super linear in $v$ uniformly in $x$ (see section 1.2-7) it follows that the sub-level set $\{\nu \in \mathcal{M}_\lambda;\ \int_{TM} l(x,v)d\nu \leq C < \infty\}$ is tight for any constant $C$, hence compact.

Next

$$\int_{TM} \left(l(x,v) - \langle d\phi, v\rangle\right) d\nu(x,v) + \int_M \phi d\lambda$$
$$= \int_M \phi d\lambda - h(x,d\phi)d\mu + \int_{TM} \left(l(x,v) - \langle d\phi, v\rangle + h(x,d\phi)\right) d\nu(x,v). \tag{4.25}$$

where $\mu = \Pi_\# \nu$. By the Young inequality $l(x,v) + h(x,\xi) \geq \langle \xi, v\rangle_{(x)}$ for any $\xi \in T^*_x M$, $v \in T_x M$ with equality if and only if $v = h_\xi(x, d\phi(x))$. So, the second term on the right of (4.25) is non-negative, but, for any $\mu \in \mathcal{M}_1^+$

$$\inf_\nu \left\{ \int_{TM} (l(x,v) - \langle d\phi, v\rangle)\, d\nu(x,v) \ \ ; \ \nu \in \mathcal{M}_1^+(TM)\ , \Pi_\# \nu = \mu \right\} = -\int_M h(x, d\phi) d\mu$$

is realized for $\nu = \delta_{v - h_\xi(x, d\phi(x))} \oplus \mu \in \mathcal{M}_1^+(TM)$. From this and (4.25) we obtain

$$\inf_{\nu \in \mathcal{M}_1^+(TM)} \mathcal{G}(\nu, \phi) = \inf_{\mu \in \mathcal{M}_1^+} \mathcal{F}(\phi, \mu)$$

hence

$$\sup_{\phi \in C^1(M)} \inf_{\nu \in \mathcal{M}_1^+(TM)} \mathcal{G}(\nu, \phi) = \sup_{\phi \in C^1(M)} \inf_{\mu \in \mathcal{M}_1^+} \mathcal{F}(\phi, \mu) = \widehat{\mathcal{C}}(\lambda)$$

and this part of the Theorem follows from (4.23).
□

## 4.4 Proof of Theorem 1:(2⥃3)

We now define, for *any* $\lambda \in \mathcal{M}_0$, a measure $\mu_\lambda \in \mathcal{M}_1^+$ in the following way:

Assume, for now, that $\lambda \in \mathcal{M}(A)$. If $E(\lambda, T) \in ]\underline{E}, \infty[ - N$ then define $\mu_\lambda = \mu_\Lambda^{E(\lambda,T)}$ according to Definition 4.1. Otherwise, fix a sequence $E^n \in ]\underline{E}, \infty[ - N$ such that $E^n \searrow E(\lambda, T)$. Similarly, let $E_n \in ]\underline{E}, \infty[ - N$ such that $E_n \nearrow E(\lambda, T)$.

Then $\mu_{\Lambda_n}^{E^n}$ and $\mu_{\Lambda_n}^{E_n}$ are given by Definition 4.1 for any $n$. Let $\mu_\lambda^+$ be a weak limit of the sequence $\mu_{\Lambda_n}^{E^n}$, and, similarly, $\mu_\lambda^-$ be a weak limit of the sequence $\mu_{\Lambda_n}^{E_n}$.

By Lemma 4.8 and Remark 4.1 we get

$$\int_M d\mu_\lambda^+ \leq T \leq \int_M d\mu_\lambda^- \ . \tag{4.26}$$

If $E(\lambda, T) = \underline{E}$ then we can still define $\mu_\lambda^+$, and it satisfies the left inequality of (4.26).

**Definition 4.3.** *For any $\lambda \in \mathcal{M}_0$, let $\mu_\lambda$ defined in the following way:*

*i) If $\lambda \in \mathcal{M}_0(A)$ then*

- *If $E(\lambda, T) > \underline{E}$ then $\mu_\lambda$ is a convex combination of $T^{-1}\mu_\lambda^+, T^{-1}\mu_\lambda^-$ given by (4.26) such that $\mu_\lambda \in \mathcal{M}_1^+$ (that is, $\int d\mu_\lambda, = 1$).*
- *If $E(\lambda, T) = \underline{E}$ then*

$$\mu_\lambda = T^{-1}\mu_\lambda^+ + \left(1 - T^{-1}\int_M d\mu_\lambda^+\right)\mu_M \tag{4.27}$$

*where $\mu_M$ is a projected Mather measure.*

*ii) For $\lambda \notin \mathcal{M}_0(A)$, let $\lambda_n \in \mathcal{M}_0(A)$ be a sequence converging weakly to $\lambda$. Then $\{\mu_\lambda\}$ is the set of weak limits of the sequence $\mu_{\lambda_n}$.*

Define

$$\mathcal{Q}(\lambda,\mu) := \sup_{\phi\in C^1(M)} \left\{ -\int_M h(x,d\phi)d\mu + \int_M \phi d\lambda \right\} \in \mathbb{R}\cup\{\infty\} \ , \quad \mathcal{Q}_T(\lambda,\mu) := \mathcal{Q}(\lambda, T\mu) \ . \tag{4.28}$$

Recall from 1⇆2 that

$$\widehat{\mathcal{C}}_T(\lambda) = \inf_{\mu\in\mathcal{M}_1^+} \mathcal{Q}_T(\lambda,\mu) \equiv \inf_{\mu\in\mathcal{M}_1^+} \mathcal{Q}(\lambda, T\mu) \ . \tag{4.29}$$

Also, from (4.13), (4.10) and Proposition 4.1

$$\overline{H}^*_T(\lambda) \le \mathcal{Q}_T(\lambda,\mu) \quad \forall \mu \in \mathcal{M}_1^+ \ . \tag{4.30}$$

We have to show that

$$\overline{H}^*_T(\lambda) = \inf_{\mu\in\mathcal{M}_1^+} \mathcal{Q}_T(\lambda,\mu) \tag{4.31}$$

for any $\lambda \in \mathcal{M}_0$. It is enough to prove (4.31) for a dense set of in $\mathcal{M}_0$, say for any $\lambda\in\mathcal{M}_0(A)$. Suppose (4.31) holds for a sequence $\{\lambda_n\} \subset \mathcal{M}_0(A)$ converging weakly to $\lambda \in \mathcal{M}_0$, that is, $\overline{H}^*_T(\lambda_n) = \widehat{\mathcal{C}}_T(\lambda_n)$. Since $\overline{H}^*_T$ is weakly continuous by Corollary 4.3 we get $\overline{H}^*_T(\lambda) = \lim_{n\to\infty} \overline{H}^*_T(\lambda_n)$. On the other hand we recall that, according to definition 2 of Theorem 1, $\widehat{\mathcal{C}}_T : \mathcal{M}_0 \mapsto \mathbb{R}$ is l.s.c. So $\lim_{n\to\infty} \widehat{\mathcal{C}}_T(\lambda_n) \ge \widehat{\mathcal{C}}_T(\lambda)$, hence $\overline{H}^*_T(\lambda) \ge \widehat{\mathcal{C}}_T(\lambda)$. By (4.29, 4.30) we get (4.31) for *any* $\lambda\in\mathcal{M}_0$.

The proof of 2 ⇆ 3 then follows from

**Lemma 4.9.** *For any $\lambda \in \mathcal{M}_0(A)$*

$$\mathcal{Q}_T(\lambda,\mu_\lambda) = \overline{H}^*_T(\lambda) \tag{4.32}$$

*holds where $\mu_\lambda \in \mathcal{M}_1^+$ is as given in Definition 4.3.*

*Proof.* Let $\lambda\in\mathcal{M}_0(A)$ and $E \in ]\underline{E},\infty[-N$. Then we use (4.21) for *any* $\phi \in C^1(M)$

$$-\int_M h(x,d\phi)d\mu^E_\Lambda = -\sum_{x,y\in A} \Lambda(\{x,y\}) \int_0^1 h\left(\boldsymbol{z}^E_{x,y}(s), d\phi\left(\boldsymbol{z}^E_{x,y}(s)\right)\right) ds$$

We now perform a change of variables $ds \to dt = \sigma'_E\left(\boldsymbol{z}^E_{x,y}(s), \dot{\boldsymbol{z}}^E_{x,y}(s)\right) ds$ which transforms the interval $[0,1]$ into $[0, T_E(x,y)]$ (see (4.7)) and we get

$$-\int_M h(x,d\phi)d\mu^E_\Lambda = -\sum_{x,y\in A} \Lambda(\{x,y\}) \int_0^{T_E(x,y)} h\left(\widehat{\boldsymbol{z}}^E_{x,y}(t), d\phi\left(\widehat{\boldsymbol{z}}^E_{x,y}(t)\right)\right) dt$$

where $\widehat{\boldsymbol{z}}^E_{x,y}$ is the re-parametrization of $\boldsymbol{z}^E_{x,y}$, satisfying $\widehat{\boldsymbol{z}}^E_{x,y}(0) = x$, $\widehat{\boldsymbol{z}}^E_{x,y}(T_E(x,y)) = y$. Next

$$\int_M \phi d\lambda = \int_M d\Lambda^E_\lambda(x,y)\left[\phi(y)-\phi(x)\right] = \sum_{x,y\in A} \Lambda(\{x,y\}) \int_0^{T_E(x,y)} d\phi\left(\widehat{\boldsymbol{z}}^E_{x,y}(t)\right) \dot{\widehat{\boldsymbol{z}}}^E_{x,y}(t) dt$$

so $\int_M \phi d\lambda - \int_M h(x,d\phi) d\mu_\Lambda^E =$

$$\sum_{x,y\in A} \Lambda_\lambda^E(\{x,y\}) \int_0^{T_E(x,y)} \left[ d\phi\left(\widehat{\boldsymbol{z}}_{x,y}^E(t)\right) \dot{\widehat{\boldsymbol{z}}}_{x,y}^E(t) - h\left(\widehat{\boldsymbol{z}}_{x,y}^E(t), d\phi\left(\widehat{\boldsymbol{z}}_{x,y}^E(t)\right)\right)\right] dt$$

$$\leq \sum_{x,y\in A} \Lambda_\lambda^E(\{x,y\}) \int_0^{T_E(x,y)} l\left(\widehat{\boldsymbol{z}}_{x,y}^E(t), \dot{\widehat{\boldsymbol{z}}}_{x,y}^E(t)\right) dt = \sum_{x,y\in A} \Lambda_\lambda^E(\{x,y\}) C_{T_E(x,y)}(x,y)$$

$$= \sum_{x,y\in A} \Lambda_\lambda^E(\{x,y\}) \left[C_{T_E(x,y)}(x,y) + ET_E(x,y)\right] - E \sum_{x,y\in A} \Lambda_\lambda^E(\{x,y\}) T_E(x,y) =$$

$$\sum_{x,y\in A} \Lambda_\lambda^E(\{x,y\}) D_E(x,y) - E \sum_{x,y\in A} \Lambda_\lambda^E(\{x,y\}) T_E(x,y) = \mathcal{D}_E(\lambda) - ED_E^{'}(\lambda) \ . \quad (4.33)$$

To obtain (4.33) we used the Young inequality in the second line, (4.8) and (4.17) on the last line.

Since (4.33) is valid for any $\phi \in C^1(M)$ we get from this and (4.30) that

$$\mathcal{D}_E(\lambda) - E\mathcal{D}_E^{'}(\lambda) \geq \mathcal{Q}(\lambda, \mu_\Lambda^E) \geq \overline{H}_T^*(\lambda) = \max_{E \geq \underline{E}} \mathcal{D}_E(\lambda) - TE \ , \quad (4.34)$$

holds for *any* $E \geq \underline{E}$. Now, if it so happens that the maximizer $E(\lambda,T)$ on the right of (4.34) is on the complement of the set $N$ in $[\underline{E},\infty[$, then $D_E^{'}(\lambda) = T = \int_M d\mu_\Lambda^E$ for $E = E(\lambda,T)$ via Lemma 4.8 and the inequality in (4.34) turns into an equality. Otherwise, if $E(\lambda,T) \in N - \{\underline{E}\}$, we take the sequences $E_n \nearrow E(\lambda,T)$, $E^n \searrow E(\lambda,T)$ for $E_n, E^n \in ]\underline{E},\infty[-N$ and the corresponding limits $\mu_\lambda^+$, $\mu_\lambda^-$ defined in (4.26). Since $\mathcal{Q}_T$ is a convex, l.s.c as a function of $\mu$ we get that the left inequality in (4.34) survives the limit, and

$$\mathcal{D}_{E(\lambda,T)}(\lambda) - E(\lambda,T) \frac{d^+}{dE} \mathcal{D}_{E(\lambda,T)}(\lambda) \geq \mathcal{Q}(\lambda, \mu_\lambda^+) \ , \ \ \mathcal{D}_{E(\lambda,T)}(\lambda) - E(\lambda,T) \frac{d^-}{dE} \mathcal{D}_{E(\lambda,T)}(\lambda) \geq \mathcal{Q}(\lambda, \mu_\lambda^-) \ , \quad (4.35)$$

while $\frac{d^+}{dE}\mathcal{D}_{E(\lambda,T)}(\lambda) = \int d\mu_\lambda^+$ and $\frac{d^-}{dE}\mathcal{D}_{E(\lambda,T)}(\lambda) = \int d\mu_\lambda^-$. Then, upon taking a convex combination $\mu_\lambda = \alpha T^{-1} \mu_\lambda^+ + T^{-1}(1-\alpha)\mu_\lambda^-$ such that, according to Definition 4.3,

$$\alpha \frac{d^+}{dE} \mathcal{D}_{E(\lambda,T)}(\lambda) + (1-\alpha) \frac{d^-}{dE} \mathcal{D}_{E(\lambda,T)}(\lambda) = T \int d\mu_\lambda = T \quad (4.36)$$

and using the convexity of $\mathcal{Q}$ in $\mu$ we get from (4.35, 4.36)

$$\mathcal{D}_{E(\lambda,T)}(\lambda) - TE(\lambda,T) \geq \mathcal{Q}(\lambda, T\mu_\lambda) \equiv \mathcal{Q}_T(\lambda, \mu_\lambda)$$

This, with the right inequality of (4.32) yields the equality $\mathcal{Q}_T(\lambda,\mu_\lambda) = \overline{H}_T^*(\lambda)$.

Finally, if $E(\lambda,T) = \underline{E}$ we proceed as follows: Let $E^n \searrow \underline{E}$ and $\mu_\lambda^+ := \lim_{n\to\infty} \mu_\lambda^{E^n}$. It follows that

$$\int_M d\mu_\lambda^+ = \lim_{n\to\infty} \int_M d\mu_\lambda^{E^n} = \lim_{n\to\infty} \mathcal{D}_{E^n}^{'}(\lambda) = \frac{d^+}{dE} \mathcal{D}_{\underline{E}}(\lambda) \in (0,T] \ . \quad (4.37)$$

Let $\mu_\lambda$ as in (4.27). From (4.28, , 4.37) and (2.4) we get

$$\mathcal{Q}_T(\lambda,\mu_\lambda) \le \mathcal{Q}(\lambda,\mu_\lambda^+) + \left(T - \frac{d^+}{dE}\mathcal{D}_{\underline{E}}(\lambda)\right)\mathcal{Q}(0,\mu_M) = \mathcal{Q}(\lambda,\mu_\lambda^+) - \left(T - \frac{d^+}{dE}\mathcal{D}_{\underline{E}}(\lambda)\right)\underline{E} \tag{4.38}$$

while (2.4) and the left part of (4.35) for $E = \underline{E}$ imply

$$\mathcal{Q}(\lambda,\mu_\lambda^+) \le \mathcal{D}_{\underline{E}}(\lambda) - \underline{E}\frac{d^+}{dE}\mathcal{D}_{\underline{E}}(\lambda)\ . \tag{4.39}$$

From (4.38) and (4.39) we get

$$\mathcal{Q}_T(\lambda,\mu_\lambda) \le \mathcal{D}_{\underline{E}}(\lambda) - \underline{E}T \le \overline{H}^*_T(\lambda)$$

and the equality holds via (4.30). The last part of Theorem 1 follows from the equality in (4.30) as well. □

### 4.5 Proof of Theorem 3

Theorem 1-(2) and (3.6) imply

$$\widehat{\mathcal{C}}_T(\lambda) = \min_{\mu\in\mathcal{M}_1^+} \widehat{\mathcal{C}}_T(\lambda\|\mu)\ . \tag{4.40}$$

Next, we note that $\mathcal{D}_E(\lambda\|\mu)$ is a concave function of $E$ for $E \ge \underline{E}$. In fact, from (3.4) and convexity of $h(x,\cdot)$ for each $x \in M$ we obtain

$$\phi_i \in \mathcal{H}_{E_i}\quad, i = 1,2 \quad\Longrightarrow \alpha\phi_1 + (1-\alpha)\phi_2 \in \mathcal{H}_{\alpha E_1 + (1-\alpha)E_2}$$

for $\alpha \in (0,1)$ and $E_1, E_2 \ge \underline{E}$. The concavity of $\mathcal{D}_{(\cdot)}(\lambda\|\mu)$ follows from its definition (3.5). Then, by convex duality and (3.6)

$$\mathcal{D}_E(\lambda\|\mu) = \min_{T>0}\left[\widehat{\mathcal{C}}_T(\lambda\|\mu) + ET\right]\ .$$

By the same argument

$$\mathcal{D}_E(\lambda) = \min_{T>0}\left[\widehat{\mathcal{C}}_T(\lambda) + ET\right]\ .$$

Hence, (4.40) and Theorem 1-(3) imply

$$\min_{\mu\in\mathcal{M}_1^+}\mathcal{D}_E(\lambda\|\mu) = \min_{\mu\in\mathcal{M}_1^+}\min_{T>0}\left[\widehat{\mathcal{C}}_T(\lambda\|\mu) + ET\right]$$

$$= \min_{T>0}\min_{\mu\in\mathcal{M}_1^+}\left[\widehat{\mathcal{C}}_T(\lambda\|\mu) + ET\right] = \min_{T>0}\left[\widehat{\mathcal{C}}_T(\lambda) + ET\right] = \mathcal{D}_E(\lambda)\ .$$

□

# 5 Proof of Theorems 2&4

## 5.1 Auxiliary results

Lemma 5.1 follows from the surjectivity of $Exp_l^{(t)}(x)$ as a mapping from $T_xM$ to $M$, for any $x \in M$ and any $t \neq 0$ (Recall definition at Section 1.2-8):

**Lemma 5.1.** *Let $\Lambda \in \mathcal{M}^+(M \times M)$. For any $t > 0$ there exists a Borel measure $\widehat{\Lambda}^{(t)} \in \mathcal{M}^+(TM)$ such that $\left(I \otimes Exp_{(l)}^{(t)}\right)_{\#} \widehat{\Lambda}^{(t)} = \Lambda$. Here $I \otimes Exp_{(l)}^{(t)}(x,v) := \left(x, Exp_{(l)}^{(t)}(x,v)\right)$.*

The proof of Lemma 5.2 follows directly from the definition of the optimal plan:

**Lemma 5.2.** *Let $\Lambda$ be a minimizer for (2.6), $B \subset M \times M$ a Borel subset and $\Lambda\lfloor_B$ the restriction of $\Lambda$ to $B$. Let $\mu_B^0$, $\mu_B^1$ the marginals of $\Lambda\lfloor_B$ on the factors of $M \times M$. Then $\Lambda\lfloor_B$ is an optimal plan for $\mathcal{C}\left(\mu_B^0, \mu_B^1\right)$. In addition, if $B_1, B_2 \subset M \times M$ are disjoint Borel sets then*

$$\mathcal{C}\left(\mu_{B_1}^0, \mu_{B_1}^1\right) + \mathcal{C}\left(\mu_{B_2}^0, \mu_{B_2}^1\right) = \mathcal{C}\left(\mu_{B_1}^0 + \mu_{B_2}^0, \mu_{B_1}^1 + \mu_{B_2}^1\right)$$

*and $\Lambda\lfloor_{B_1 \cup B_2}$ is the optimal plan with respect to $\mathcal{C}\left(\mu_{B_1}^0 + \mu_{B_2}^0, \mu_{B_1}^1 + \mu_{B_2}^1\right)$.*

Lemma 5.3 represents the *time interpolation* of optimal plans (see [28]):

**Lemma 5.3.** *Given $t > 0$ and $\lambda = \lambda^+ - \lambda^- \in \mathcal{M}_0$. Let $\Lambda^t \in \mathcal{P}(\lambda^+, \lambda^-)$ be an optimal plan realizing*

$$\mathcal{C}_t(\lambda^+, \lambda^-) = \int\int C_t(x,y)\Lambda^t(dxdy) \ .$$

*Let $\widehat{\Lambda}^{(t)} \in \mathcal{M}^+(TM)$ given in Lemma 5.1 for $\Lambda = \Lambda^t$. Let $\lambda_s := \left(Exp_l^{(s)}\right)_{\#} \widehat{\Lambda}^{(t)}$. Then, if $0 < s < t$,*

$$\mathcal{C}_s(\lambda^+, \lambda_s) + \mathcal{C}_{t-s}(\lambda_s, \lambda^-) = \mathcal{C}_t(\lambda^+, \lambda^-) \ .$$

**Lemma 5.4.** *For any $\lambda^+, \lambda^- \in \mathcal{M}_1^+$ satisfying $\lambda = \lambda^+ - \lambda^- \in \mathcal{M}_1^+$,*

$$\mathcal{C}_T(\lambda^+, \lambda^-) \geq \widehat{\mathcal{C}}_T(\lambda) \ .$$

**Lemma 5.5.** *$\widehat{\mathcal{C}}_T(\lambda \| \mu)$ is l.s.c in the weak-* topology of $\mathcal{M}_0 \times \mathcal{M}_1^+$. Assuming $\mathbf{H_1}$ and $\mathbf{H_2}$, for any $\lambda \in \mathcal{M}_0$, $\mu \in \mathcal{M}_1^+$ there exists a sequence $\{\tilde{\mu}_n\} = \{\rho_n(x)dx\} \subset \mathcal{M}_1^+$, $\{\tilde{\lambda}_n\} = \{\rho_n(q_n^+ - q_n^-)dx\} \subset \mathcal{M}_0$ where $\rho_n \in C^\infty(M)$ are positive everywhere, $q_n^\pm \in C^\infty(M)$ non-negatives such that $\tilde{\lambda}_n \rightharpoonup \lambda$, $\tilde{\mu}_n \rightharpoonup \mu$ and*

$$\lim_{n\to\infty} \widehat{\mathcal{C}}_T(\tilde{\lambda}_n \| \tilde{\mu}_n) = \widehat{\mathcal{C}}_T(\lambda \| \mu) \ . \tag{5.1}$$

**Lemma 5.6.** *For any $\mu \in \mathcal{M}_1^+$, $\lambda = \lambda^+ - \lambda^- \in \mathcal{M}_0$*

$$\liminf_{\varepsilon\to 0} \varepsilon^{-1}\mathcal{C}_{\varepsilon T}(\mu + \varepsilon\lambda^+, \mu + \varepsilon\lambda^-) \geq \widehat{\mathcal{C}}_T(\lambda \| \mu) \ .$$

**Lemma 5.7.** *Assume $\mu = \rho(x)dx$ and $\lambda = \rho(q^+ - q^-)dx$ where $\rho, q^\pm$ are $C^\infty$ functions, $\rho$ positive everywhere on $M$. Then*

$$\limsup_{\varepsilon\to 0} \varepsilon^{-1}\mathcal{C}_{\varepsilon T}\left(\mu + \varepsilon\lambda^+, \mu + \varepsilon\lambda^-\right) \leq \widehat{\mathcal{C}}_T(\lambda \| \mu) \ .$$

**Lemma 5.8.** *For $T > 0$,*

$$\widehat{\mathcal{C}}_T(\lambda) \geq \limsup_{\varepsilon\to 0} \varepsilon^{-1} \inf_{\mu\in\mathcal{M}_1^+} \mathcal{C}_{\varepsilon T}(\mu+\varepsilon\lambda^+, \mu+\varepsilon\lambda^-) \ .$$

*Proof. of Lemma 5.4*: We use the duality representation of the Monge-Kantorovich functional [26] to obtain (recall $\lambda^{\pm} \in \mathcal{M}_1^+$)

$$\mathcal{C}_T(\lambda^+,\lambda^-) + ET = \sup_{\psi,\phi}\left\{ \int_M \psi d\lambda^- - \phi d\lambda^+ \quad , \quad \phi(y)-\psi(x) \leq C_T(x,y)+ET \right\}$$

By (2.10) $C_T(x,y) + ET \geq D_E(x,y)$ for any $x, y \in M$ so, by (2.12, 2.13)

$$\sup_{\psi,\phi}\left\{ \int_M \psi d\lambda^- - \phi d\lambda^+ \ , \ \phi(y)-\psi(x) \leq C_T(x,y)+ET \right\} \geq \sup_{\phi}\left\{ \int_M \phi d\lambda \ , \ \phi(y)-\phi(x) \leq D_E(x,y) \right\}$$
$$= \mathcal{D}_E(\lambda) \quad (5.2)$$

so

$$\mathcal{C}_T(\lambda^+,\lambda^-) \geq \mathcal{D}_E(\lambda) - ET$$

for *any* $E \geq \underline{E}$. By Theorem 1-(3)

$$\mathcal{C}_T(\lambda^+,\lambda^-) \geq \sup_{E\geq \underline{E}} \mathcal{D}_E(\lambda) - ET = \widehat{\mathcal{C}}_T(\lambda) \ .$$

□

*Proof. of Lemma 5.5*: From (3.5, 3.6) we obtain

$$\widehat{\mathcal{C}}_T(\lambda\|\mu) = \sup_{\phi\in C^1(M)} \int_M \phi d\lambda - Th(x,d\phi)d\mu \ .$$

In particular $\widehat{\mathcal{C}}_T$ is l.s.c (and convex) on $\mathcal{M}_0 \times \mathcal{M}_1^+$.

Let $\varepsilon_n \to 0$ and $\lambda_n := \lambda_{\varepsilon_n} := \delta_{\varepsilon_n} * \lambda \in \mathcal{M}_0$ defined by

$$\int_M \psi d\lambda_n := \lambda(\delta_{\varepsilon_n} * \psi) \quad \forall \psi \in C^0(M) \ . \tag{5.3}$$

By $\mathbf{H_1}$, $\lambda_n \rightharpoonup \lambda$ while $\lambda_n$ are smooth. First, we observe that $\lim_{n\to\infty}\lambda_n \rightharpoonup \lambda$. Indeed, for any $\psi \in C^1(M)$:

$$\lim_{n\to\infty} \int_M \psi d\lambda_n = \lim_{n\to\infty} \lambda\left(\delta_{\varepsilon_n} * \psi\right) = \lambda(\psi) \ .$$

Next, by Jensen's Theorem and $\mathbf{H_2}$

$$\int_M h(x, d\delta_\varepsilon * \phi)d\mu = \int_M h(x,\delta_\varepsilon * d\phi)d\mu \leq \int_{M\times M} h(x,d\phi(y))\delta_\varepsilon(x,y)d\mu(x)dy$$
$$\equiv \int_M h(x,d\phi)d\delta_\varepsilon * \mu + \int_{M\times M} \left[h(x,d\phi(y)) - h(y,d\phi(y)\right]\delta_\varepsilon(x,y)d\mu(x)dy \quad (5.4)$$

From section 1.2-(7) and using $\delta_\varepsilon(x,y)=o(1)$ for $D(x,y)>\delta$,

$$\int_{M\times M}[h(x,d\phi(y))-h(y,d\phi(y)]\,\delta_\varepsilon(x,y)d\mu(x)dy\leq O(\varepsilon)+o(1)\int_M h(x,d\phi)d\delta_\varepsilon*\mu\ .$$

Next, define $\mu_n=\delta_{\varepsilon_n}*\mu$. Let $\psi_n$ be the maximizer of $\widehat{\mathcal{C}}(\lambda_n\|\mu_n)$, that is

$$\widehat{\mathcal{C}}_T(\lambda_n\|\mu_n)=\int_M\psi_n d\lambda_n-Th(x,d\psi_n)d\mu_n$$

By (5.3, 5.4)

$$\widehat{\mathcal{C}}_T(\lambda_n\|\mu_n)\leq\int_M\delta_\varepsilon*\psi_n d\lambda-(1-o(1))\int_M Th(x,d\delta_\varepsilon*\psi_n)d\mu+O(\varepsilon_n)=$$
$$(1-o(1))\left[\int_M\delta_\varepsilon*\psi_n\frac{d\lambda}{1-o(1)}-\int_M Th(x,d\delta_\varepsilon*\psi_n)d\mu\right]+\varepsilon_n\leq(1-o(1))\widehat{\mathcal{C}}\left(\frac{\lambda}{1-o(1)}\|\mu\right)+\varepsilon_n \tag{5.5}$$

We obtained

$$\limsup_{n\to\infty}\widehat{\mathcal{C}}_T(\lambda_n\|\mu_n)\leq\widehat{\mathcal{C}}_T(\lambda\|\mu)$$

which, together with the l.s.c of $\widehat{\mathcal{C}}_T$, implies the result. □

*Proof. of Lemma 5.6*: Recall that the Lax-Oleinik Semigroup acting on $\phi\in C^0(M)$

$$\psi(x,t)=LO(\phi)_{(t,x)}:=\sup_{y\in M}[\phi(y)-C_t(x,y)]$$

is a viscosity solution of the Hamilton-Jacobi equation $\partial_t\psi-h(x,d\psi)=0$ subjected to $\psi_0=\phi(x)$. If $\phi\in C^1(M)$ then $\psi$ is a *classical solution* on some neighborhood of $t=0$, so

$$\lim_{T\to 0}LO(\phi)_{(T,\cdot)}=\phi\ \ ;\ \ \lim_{T\to 0}T^{-1}\left[LO(\phi)_{(T,x)}-\phi(x)\right]=h(x,d\phi)\ .$$

Then for any $\mu_1,\mu_2\in\mathcal{M}_1^+$

$$\mathcal{C}_T(\mu_1,\mu_2)=\sup_{\phi,\psi\in C^1(M)}\left\{\int_M\phi d\mu_2-\psi d\mu_1\quad;\quad\phi(x)-\psi(y)\leq C_T(x,y)\ \ \forall x,y\in M\right\}=$$
$$\sup_{\phi\in C^1(M)}\int_M\phi d\mu_2-LO(\phi)_{(T,x)}d\mu_1 \tag{5.6}$$

Hence

$$\liminf_{\varepsilon\to 0}\varepsilon^{-1}\mathcal{C}_{\varepsilon T}(\mu+\varepsilon\lambda^+,\mu+\varepsilon\lambda^-)=$$
$$\liminf_{\varepsilon\to 0}\sup_{\phi\in C^1(M)}\int_M\varepsilon^{-1}\left[\phi(x)-LO(\phi)_{(\varepsilon T,x)}\right]d\mu+\int_M\phi d\lambda^+-LO(\phi)_{(\varepsilon T,x)}d\lambda^-$$
$$\geq\sup_{\phi\in C^1(M)}\lim_{\varepsilon\to 0}\int_M\varepsilon^{-1}\left[\phi(x)-LO(\phi)_{(\varepsilon T,x)}\right]d\mu+\int_M\phi d\lambda^+-LO(\phi)_{(\varepsilon T,x)}d\lambda^-$$
$$=\sup_{\phi,\psi\in C^1(M)}\int_M-Th(x,d\phi)d\mu+\phi d\lambda:=\widehat{\mathcal{C}}_T(\lambda\|\mu)\ . \tag{5.7}$$

□

*Proof. of Lemma 5.7*: We may describe the optimal mapping $S_{\varepsilon T}: M \to M$ associated with $C_{\varepsilon T}(\mu+\varepsilon\lambda^+, \mu+\varepsilon\lambda^-)$ in local coordinates on each chart. It is given by the solution to the Monge-Ampère equation

$$det\nabla_x S_{\varepsilon T} = \frac{\rho(x)(1+\varepsilon q^-(x))}{\rho(S_{\varepsilon T}(x))(1+\varepsilon T q^+(S_{\varepsilon T}(x))} \tag{5.8}$$

where

$$\nabla\psi = -\nabla_x C_{\varepsilon T}(x, S_{\varepsilon T}(x)) \tag{5.9}$$

and

$$\mathcal{C}_{\varepsilon T}\left(\mu+\varepsilon\lambda^+, \mu+\varepsilon\lambda^-\right) = \int_M C_{\varepsilon T}(x, S_{\varepsilon T}(x))\rho(1+\varepsilon T q^-)dx \tag{5.10}$$

We recall that the inverse of $\nabla_x C_{\varepsilon T}(x,\cdot)$ with respect to the second variable is $I_d + \varepsilon T\nabla\psi$, to leading order in $\varepsilon$. That is,

$$\nabla_x C_{\varepsilon T}\left(x, x+\varepsilon T\partial_p h(x,\xi) + (\varepsilon T)^2 Q(x,\xi,\varepsilon)\right) = -\xi \tag{5.11}$$

where (here and below) $Q$ is a generic smooth function of its arguments.

Hence, $S_{\varepsilon T}$ can be expanded in $\varepsilon$ in terms of $\psi$ as

$$S_{\varepsilon T}(x) = x + \varepsilon T h_\xi(x,\nabla\psi) + (\varepsilon T)^2 Q(x,\nabla\psi,\varepsilon) \tag{5.12}$$

We now expand the right side of (5.8) using (5.12) to obtain

$$1+\varepsilon T\left[q^-(x) - q^+(x) - h_\xi(x,d\psi)\cdot\nabla_x \ln\rho(x)\right] + (\varepsilon T)^2 Q(x,\nabla\psi,x,\varepsilon) \tag{5.13}$$

while the left hand side is

$$det(\nabla_x S_{\varepsilon T}) = 1+\varepsilon T\nabla\cdot h_\xi(x,d\psi) + (\varepsilon T)^2 Q(x,\nabla\psi,\nabla\nabla\psi,x,\varepsilon) \tag{5.14}$$

Comparing (5.13, 5.14), divide by $\varepsilon T$ and multiply by $\rho$ to obtain

$$T\nabla\cdot(\rho h_\xi(x,d\psi)) = \rho(q^- - q^+) + \varepsilon T\rho Q(x,\nabla\psi,\nabla\nabla\psi,x,\varepsilon)\ . \tag{5.15}$$

Now, we substitute $\varepsilon = 0$ and get a quasi-linear equation for $\psi_0$:

$$T\nabla\cdot(\rho h_\xi(x,d\psi_0)) = \rho(q^- - q^+)\ . \tag{5.16}$$

$\psi_0$ is a maximizer of

$$\widehat{\mathcal{C}}_T(\lambda\|\mu) = \int_M \rho(q^+ - q^-)\psi_0 - \int_M \rho T h(x,d\psi_0)dx$$

By elliptic regularity, $\psi_0 \in C^\infty(M)$. Multiply (5.16) by $\psi_0$ and integrate over $M$ to obtain

$$\int_M \rho(q^+ - q^-) = \int_M \rho T h_\xi(x,d\psi_0)\cdot\nabla\psi_0$$

Then by the Lagrangian/Hamiltonian duality

$$\widehat{\mathcal{C}}_T(\lambda\|\mu) = \int_M \rho T\left[\nabla\psi_0 \cdot h_\xi(x, d\psi_0) - h(x, d\psi_0)\right] \equiv T\int_M \rho l\left(x, h_\xi(x, d\psi_0)\right) \ . \tag{5.17}$$

We observe $l\left(x, \frac{y-x}{T}\right) \geq T^{-1}C_T(x,y)$. So, (5.10) with (5.12) imply

$$(\varepsilon T)^{-1}\mathcal{C}_{\varepsilon T}\left(\mu+\varepsilon\lambda^+, \mu+\varepsilon\lambda^-\right) \leq \int_M \rho(1+\varepsilon T q^-)l\left(x, h_\xi(x, \nabla\psi_\varepsilon + \varepsilon T Q(x, \nabla\psi_\varepsilon, \varepsilon)\right) \tag{5.18}$$

where $\psi_\varepsilon$ is a solution of (5.15). Now, if we show that $\lim_{\varepsilon\to 0}\psi_\varepsilon = \psi_0$ in $C^1(M)$ then, from (5.17, 5.18)

$$\limsup_{\varepsilon\to 0}(\varepsilon)^{-1}\mathcal{C}_{\varepsilon T}\left(\mu+\varepsilon\lambda^+, \mu+\varepsilon\lambda^-\right) \leq T\int_M \rho l\left(x, h_\xi(x, d\psi_0)\right) = \widehat{\mathcal{C}}(\lambda\|\mu) \ .$$

Next we show that, indeed, $\lim_{\varepsilon\to 0}\psi_\varepsilon = \psi_0$ in $C^1(M)$.

Substitute $\psi_\varepsilon = \psi_0 + \phi_\varepsilon$ in (5.15). We obtain

$$\nabla\cdot(\sigma(x)\nabla\phi_\varepsilon) = \varepsilon Q(x, \nabla\phi_\varepsilon, \nabla\nabla\phi_\varepsilon, \varepsilon) + \nabla\cdot\left(\rho\langle\nabla^t\phi_\varepsilon, \tilde{Q}(x, \nabla\phi, \varepsilon)\cdot\nabla\phi_\varepsilon\rangle\right) \tag{5.19}$$

where $\sigma := Th_{\xi\xi}(x, \nabla\psi_0(x))$ is a positive definite form, while $\tilde{Q}$ is a smooth matrix valued functions in both $x$ and $\varepsilon$, determined by $\nabla\psi_0$ and $Q$ as given in (5.15). A direct application of the implicit function theorem implies the existence of a branch $(\lambda(\varepsilon), \eta_\varepsilon)$ of solutions for

$$\nabla\cdot(\sigma(x)\nabla\eta) = \varepsilon Q(x, \nabla\eta, \nabla\nabla\eta, \varepsilon) + \nabla\cdot\left(\rho\langle\nabla^t\eta, \tilde{Q}(x, \nabla\eta, \varepsilon)\circ\nabla\eta\rangle\right) + \lambda(\varepsilon) \tag{5.20}$$

where $\eta_0 = \lambda(0) = 0$ and $\varepsilon\mapsto\eta_\varepsilon$ is (at least) continuous in $C^1(M)\perp 1$. Note that for $\varepsilon\neq 0$ we may have a non-zero $\lambda(\varepsilon)$ which follows from projecting the right side on the equation to the Hilbert space perpendicular to constants (recall that $M$ is a compact manifold without boundary, and the left side is surjective on this space). We now show that $\eta_\varepsilon = \phi_\varepsilon$, i.e $\lambda(\varepsilon) = 0$ also for $\varepsilon\neq 0$. Indeed, (5.19) is equivalent to (5.8) multiplied by $\rho(x)/\varepsilon$, so (5.20) is equivalent to

$$det\nabla_x\hat{S}_{\varepsilon T} = \frac{\rho(x)(1+\varepsilon q^-(x))}{\rho(\hat{S}_{\varepsilon T}(x))(1+\varepsilon q^+(\hat{S}_{\varepsilon T}(x))} + \varepsilon\rho^{-1}(x)\lambda(\varepsilon)$$

where $\hat{S}_{\varepsilon T}(x)$ obtained from (5.12) with $\psi_\varepsilon := \psi_0 + \eta_\varepsilon$.

Hence

$$\int_M \left(\rho(\hat{S}_{\varepsilon T}(x))(1+\varepsilon q^+(\hat{S}_{\varepsilon T}(x))\right) det(\nabla_x\hat{S}_{\varepsilon T}) = \int_M \left(\rho(x)(1+\varepsilon q^-(x))\right) + \varepsilon\lambda(\varepsilon)\int_M \frac{\rho(\hat{S}_{\varepsilon T}(x))}{\rho(x)}(1+\varepsilon q^+(\hat{S}_{\varepsilon T}(x)) \tag{5.21}$$

However, $\hat{S}_{\varepsilon T}(x) = x + O(\varepsilon)$ is a diffeomorphism on $M$, so

$$\int_M \left(\rho(\hat{S}_{\varepsilon T}(x))(1+\varepsilon q^+(\hat{S}_{\varepsilon T}(x))\right) det(\nabla_x \hat{S}_{\varepsilon T}) = \int_M \left(\rho(\hat{S}_{\varepsilon T}(x))(1+Tq^+(\hat{S}_{\varepsilon T}(x))\right) |det(\nabla_x \hat{S}_{\varepsilon T})|$$
$$= \int_M \rho(x)(1+\varepsilon q^+(x)) \equiv \int_M \rho(x)(1+\varepsilon q^-(x)) \ . \quad (5.22)$$

It follows that

$$\varepsilon\lambda(\varepsilon)\int_M \frac{\rho(\hat{S}_{\varepsilon T}(x))}{\rho(x)}(1+\varepsilon q^+(\hat{S}_{\varepsilon T}(x)) = 0 \ .$$

Since $\rho$ is positive everywhere it follows that $\lambda(\varepsilon) \equiv 0$ for $|\varepsilon|$ sufficiently small. We proved that $\eta_\varepsilon \equiv \phi_\varepsilon$ and, in particular, $\phi_\varepsilon \to 0$ as $\varepsilon \to 0$ in $C^1 \perp 1$, which implies the convergence of $\psi_\varepsilon$ to $\psi_0$ at $\varepsilon \to 0$ in $C^1 \perp 1$. □

*Proof.* (of Lemma 5.8) Given $\varepsilon > 0$ let

$$D^\varepsilon_E(x,y) := \inf_{n\in\mathbb{N}} \left[C_{\varepsilon n T}(x,y) + \varepsilon n E T\right] \ . \quad (5.23)$$

Evidently, $D^\varepsilon_E(x,y)$ is continuous on $M \times M$ locally uniformly in $E \geq \underline{E}$. Moreover,

$$\lim_{\varepsilon \searrow 0} D^\varepsilon_E = D_E \quad (5.24)$$

uniformly on $M \times M$ and locally uniformly in $E \geq \underline{E}$ as well.

We now decompose $M \times M$ into mutually disjoint Borel sets $Q_n$:

$$M \times M = \cup_n Q^\varepsilon_n \ , \quad Q^\varepsilon_n \cap Q^\varepsilon_{E,n'} = \emptyset \ \text{ if } n \neq n^{'}$$

such that

$$Q^\varepsilon_n \subset \{(x,y) \in M \times M \ ; \ \ D^\varepsilon_E(x,y) = C_{\varepsilon n T}(x,y) + \varepsilon n E T\} \ .$$

Let $\Lambda^E_\varepsilon \in \mathcal{P}(\lambda^+, \lambda^-)$ be an optimal plan for

$$\mathcal{D}^\varepsilon_E(\lambda) = \int_{M\times M} D^\varepsilon_E(x,y) d\Lambda^E_\varepsilon = \min_{\Lambda\in\mathcal{P}(\lambda^+,\lambda^-)} \int_{M\times M} D^\varepsilon_E(x,y) d\Lambda \ , \quad (5.25)$$

and $\Lambda^n_\varepsilon = \Lambda^E_\varepsilon \lfloor_{Q^\varepsilon_n}$, the restriction of $\Lambda^E_\varepsilon$ to $Q^\varepsilon_n$. Set $\lambda^\pm_n$ to be the marginals of $\Lambda^n_\varepsilon$ on the first and second factors of $M \times M$. Then $\sum_{n=1}^\infty \Lambda^n_\varepsilon = \Lambda^E_\varepsilon$ and

$$\sum_{n=1}^\infty \lambda^\pm_n = \lambda^\pm \quad (5.26)$$

**Remark 5.1.** *Note that $Q^\varepsilon_n = \emptyset$ for all but a finite number of $n \in \mathbb{N}$. In particular, the sum (5.26) contains only a finite number of non-zero terms.*

Let $|\lambda_n| := \int_M d\lambda_n^\pm \equiv \int_{M\times M} d\Lambda^\varepsilon_n$. The *averaged flight time* is

$$\langle T\rangle^\varepsilon := \varepsilon T \sum_{n=1}^\infty n|\lambda_n| \tag{5.27}$$

We observe that $\langle T\rangle^\varepsilon \in \partial_E \mathcal{D}^\varepsilon_E(\lambda)$, where $\partial_E$ is the super gradient as a function of $E$. At this stage we choose $E$ depending on $\varepsilon, T$ such that

$$\langle T\rangle^\varepsilon = T + 2\varepsilon T|\lambda^\pm| \tag{5.28}$$

We now apply Lemma 5.1: Recalling Section 1.2-8, let $\widehat{\Lambda}^n_\varepsilon \in \mathcal{M}^+(TM)$ satisfying $\left(I \oplus Exp_{(l)}^{(t=\varepsilon nT)}\right)_\# \widehat{\Lambda}^n_\varepsilon = \Lambda^n_\varepsilon$. Use $\widehat{\Lambda}^n_\varepsilon$ to define $\lambda^j_n := \left(Exp_{(l)}^{(t=\varepsilon nT)}\right)_\# \widehat{\Lambda}^n_\varepsilon \in \mathcal{M}^+(M)$ for $j = 0, 1 \dots n$. Note that

$$\lambda^0_n = \lambda^+_n \quad , \lambda^n_n = \lambda^-_n \ . \tag{5.29}$$

By Lemma 5.3

$$\mathcal{C}_{\varepsilon nT}(\lambda^+_n, \lambda^-_n) + \varepsilon nET|\lambda_n| = \sum_{j=0}^{n-1}\left[\mathcal{C}_{\varepsilon T}(\lambda^j_n, \lambda^{j+1}_n) + \varepsilon ET|\lambda_n|\right] \tag{5.30}$$

From (5.23, 5.25, 5.26, 5.30) and Lemma 5.2

$$\mathcal{D}^\varepsilon_E(\lambda) = \sum_{n=1}^\infty \mathcal{D}^\varepsilon_E(\lambda_n) = \sum_{n=1}^\infty \left[\mathcal{C}_{\varepsilon nT}(\lambda^+_n, \lambda^-_n) + \varepsilon nET|\lambda_n|\right] = \sum_{n=1}^\infty\sum_{j=0}^{n-1}\left(\mathcal{C}_{\varepsilon T}(\lambda^j_n, \lambda^{j+1}_n) + \varepsilon ET|\lambda_n|\right) \ . \tag{5.31}$$

Let now

$$\mu^{\varepsilon,E} = \varepsilon \sum_{n=1}^\infty\sum_{j=1}^{n-1} \lambda^j_n \ .$$

Note that

$$\mu^{\varepsilon,E} = \varepsilon \sum_{n=1}^\infty\sum_{j=0}^{n} \lambda^j_n - \varepsilon\sum_{n=1}^\infty \lambda^0_n - \varepsilon \sum_{n=1}^\infty \lambda^n_n \ .$$

By (5.26,5.29, 5.27) we obtain

$$\left|\mu^{\varepsilon,E}\right| = \varepsilon \sum_{n=1}^\infty (n+1)|\lambda^\pm_n| - 2\varepsilon|\lambda^\pm| = 1 \implies \mu^{\varepsilon,E} \in \mathcal{M}^+_1 \ . \tag{5.32}$$

By (5.26, 5.29)

$$\sum_{n=1}^\infty\sum_{j=0}^{n-1} \mathcal{C}_{\varepsilon T}(\lambda^j_n, \lambda^{j+1}_n) \geq \mathcal{C}_{\varepsilon T}\left(\sum_{n=1}^\infty\sum_{j=0}^{n-1}\lambda^j_n, \sum_{n=1}^\infty\sum_{j=1}^{n}\lambda^{j+1}_n\right) = \varepsilon^{-1}\mathcal{C}_{\varepsilon T}\left(\varepsilon\sum_{n=1}^\infty\sum_{j=0}^{n-1}\lambda^j_n, \varepsilon\sum_{n=1}^\infty\sum_{j=1}^{n}\lambda^{j+1}_n\right)$$
$$= \varepsilon^{-1}\mathcal{C}_{\varepsilon T}\left(\mu^{\varepsilon,E} + \varepsilon\lambda^+, \mu^{\varepsilon,E} + \varepsilon\lambda^-\right) \ . \tag{5.33}$$

From (5.27, 5.31, 5.33 , 5.32)

$$\mathcal{D}_E^{\varepsilon}(\lambda) - \langle T\rangle^{\varepsilon} E \geq \varepsilon^{-1}\mathcal{C}_{\varepsilon T}\left(\mu^{\varepsilon,E} + \varepsilon\lambda^+, \mu^{\varepsilon,E} + \varepsilon\lambda^-\right) \geq \varepsilon^{-1} \inf_{\mu\in\mathcal{M}_1^+} \mathcal{C}_{\varepsilon T}\left(\mu + \varepsilon\lambda^+, \mu + \varepsilon\lambda^-\right) \ . \tag{5.34}$$

Finally, Theorem 1-3, (5.24, 5.28, 5.34) imply

$$\widehat{\mathcal{C}}_T(\lambda) \geq \mathcal{D}_E(\lambda) - TE = \lim_{\varepsilon\to 0} \mathcal{D}_E^{\varepsilon}(\lambda) - \langle T\rangle^{\varepsilon} E \geq \limsup_{\varepsilon\to 0} \varepsilon^{-1} \inf_{\mu\in\mathcal{M}_1^+} \mathcal{C}_{\varepsilon T}\left(\mu + \varepsilon\lambda^+, \mu + \varepsilon\lambda^-\right) \ .$$

□

### 5.2 Proof of theorem 2

From Theorem 1- (1) we get

$$\widehat{\mathcal{C}}_{\varepsilon T}(\varepsilon\lambda) = \varepsilon\widehat{\mathcal{C}}_T(\lambda) \ .$$

We now apply Lemma 5.4, adapted to the case where $|\lambda^{\pm}| := \int \lambda^{\pm} \neq 1$. Then

$$\mathcal{C}_T(\lambda^+, \lambda^-) = |\lambda^{\pm}|\mathcal{C}_T\left(\frac{\lambda^+}{|\lambda^+|}, \frac{\lambda^-}{|\lambda^-|}\right) \geq |\lambda^{\pm}|\widehat{\mathcal{C}}_T\left(\frac{\lambda}{|\lambda^{\pm}|}\right) = \widehat{\mathcal{C}}_{T/|\lambda^{\pm}|}(\lambda) \ .$$

Note that $\int d\mu + \varepsilon d\lambda^{\pm} = 1 + O(\varepsilon)$, hence

$$\varepsilon^{-1}\mathcal{C}_{\varepsilon T}\left(\mu + \varepsilon\lambda^+, \mu + \varepsilon\lambda^-\right) \geq \widehat{\mathcal{C}}_{T_\varepsilon}(\lambda)$$

where $T_\varepsilon \to T$ as $\varepsilon \to 0$. Hence

$$\liminf_{\varepsilon\to 0} \inf_{\mathcal{M}_1^+} \varepsilon^{-1}\mathcal{C}_{\varepsilon T}\left(\mu + \varepsilon\lambda^+, \mu + \varepsilon\lambda^-\right) \geq \widehat{\mathcal{C}}_T(\lambda) \ .$$

The Theorem follows from this and Lemma 5.8.

□

### 5.3 Proof of Theorem 4

We have to show that for any $(\mu, \lambda) \in \mathcal{M}_1^+ \times \mathcal{M}_0$ and any sequence $(\mu_n, \lambda_n) \rightharpoonup (\mu, \lambda)$ as $n \to \infty$:

$$\liminf_{n\to\infty} n\mathcal{C}_{T/n}\left(\mu_n + n^{-1}\lambda_n^+, \mu_n + n^{-1}\lambda_n^-\right) \geq \widehat{\mathcal{C}}(\lambda\|\mu) \tag{5.35}$$

and, in addition, *there exists* a sequence $(\hat{\mu}_n, \hat{\lambda}_n) \rightharpoonup (\mu, \lambda)$ for which

$$\lim_{n\to\infty} n\mathcal{C}_{T/n}\left(\hat{\mu}_n + n^{-1}\hat{\lambda}_n^+, \hat{\mu}_n + n^{-1}\hat{\lambda}_n^-\right) = \widehat{\mathcal{C}}(\lambda\|\mu) \ . \tag{5.36}$$

The inequality (5.35) follows directly from Lemma 5.6. To prove (5.36), we first consider the sequence $(\tilde{\mu}_n, \tilde{\lambda}_n)$ subjected to Lemma 5.5. From Lemma 5.7 and Lemma 5.5,

$$\lim_{j\to\infty} \limsup_{n\to\infty} \ n\mathcal{C}_{T/n}\left(\tilde{\mu}_j + n^{-1}\tilde{\lambda}_j^+, \tilde{\mu}_j + n^{-1}\tilde{\lambda}_j^-\right) \leq \lim_{j\to\infty} \widehat{\mathcal{C}}_T\left(\tilde{\lambda}_j\|\tilde{\mu}_j\right) = \widehat{\mathcal{C}}(\lambda\|\mu) \ .$$

So, there exists a subsequence $j_n$ along which

$$\limsup_{n\to\infty} \ n\mathcal{C}_{T/n}\left(\tilde{\mu}_{j_n} + n^{-1}\tilde{\lambda}^{+}_{j_n}, \tilde{\mu}_{j_n} + n^{-1}\tilde{\lambda}^{-}_{j_n}\right) \le \widehat{\mathcal{C}}(\lambda\|\mu) \ .$$

This, with (5.35), implies (5.36).

The second part of the theorem follows from (5.35) and Theorem 2.

□

## References


[1] Ambrosio, L and Pratelli, L: *Existence and stability results in the L1 theory of optimal transportation*, Lect. Notes in Math, **1813**, (2003)

[2] Bangert, V: *Minimal measures and minimizing closed normal one-currents*, GAFA, **9**, 413-427, (1999)

[3] Bernard, P: *Young measures, superposition and transport*, Indiana Univ. Math. J. **57** # 1, 247-275, (2008)

[4] Bernard P. and Buffoni B: *Optimal mass transportation and Mather theory* , J. Eur. Math. Soc. **9**, 85121, (2007)

[5] Bernard P. and Buffoni B: *Weak KAM pairs and Monge-Kantorovich duality*, Adv. Study in Pure Mathematics, **47**, 397-420, (2007)

[6] Bouchitt, G., Buttazzo, G. and Seppecher, P.: *Energies with respect to a measure and applications to low dimensional structures.* Calc. Var. Partial Differential Equations **5**, 3754, (1997)

[7] Bouchitt, G., Buttazzo, G. and Seppecher, P.: *Shape optimization solutions via MongeKantorovich equation*, CRAS 324 , 1185- 1191. (1997)

[8] Buttazzo ,G. and Stepanov, E.: *Optimal transportation networks as free Dirichlet regions for the Monge-Kantorovich problem.* Ann. Scuola Norm. Sup. Pisa Cl. Sci., (5) **2** , 631-678, (2003)

[9] Bernot, M, Caselles, V and Morel, J.M.: *Optimal Transportation Networks*, Lect. Notes in Math., **1955**, Springer-Verlag (2009)

[10] G. Contreras, J. Delgado and R. Iturriaga: *Lagrangian flows: the dynamics of globally minimizing orbits. II*, Bol. Soc. Brasil. Mat. (N.S.) **28** , # 2, 155-196, (1997)

[11] Evans, L.C: *A survey of partial differential equations methods in weak KAM theory*, Comm. Pure Appl. Math. **57** 4, 445-480, (2004)

[12] Evans L.C. and Gangbo W: *Differential equations methods for the Monge-Kantorovich mass transfer problem*, Mem. Amer. Math. Soc. **137**, (1999)

[13] Evans L.C and D. Gomes, D: *Effective Hamiltonians and averaging for Hamiltonian dynamics. I.*, Arch. Ration. Mech. Anal. **157** 1, 1-33, (2001)

[14] Evans L.C and D. Gomes, D: *Effective Hamiltonians and averaging for Hamiltonian dynamics. II.* Arch. Ration. Mech. Anal. **161** , 4, 271-305, (2002)

[15] Fathi A: *Weak KAM Theorem in Lagrangian Dynamics*, Cambridge Studies in Advanced Mathematics (in press)

[16] Fathi A: *Solutions KAM faibles conjuguées et barrières de Peierls*, C. R. Acad. Sci. Paris Sèr. I Math. 325 , 649652 (1997)

[17] Fathi A. and Sicololfi A: *PDE aspects of Aubry-Mather theory for quasiconvex hamiltonians*, Cal. Var. **22**, 185-228 (2005)

[18] Fathi, A and Siconolfi, A: *Existence of $C^1$ critical subsolutions of the Hamilton-Jacobi equation.* Invent. Math. **155** , 2, 363-388, (2004)

[19] Iri, M: *Theory of flows in continua as approximation to flows in networks*, in Survey of Math Programming (ed. by Prekopa), North-Holland, 1979.

[20] Mather J.N: *Action minimizing invariant measures for positive definite Lagrangian systems*, Math. Z. **207** , 169-207 (1991)

[21] Mañé R: *On the minimizing measures of Lagrangian dynamical systems*, Nonlinearity **5**, 623-638, (1992)

[22] Mañé R: *Lagrangian flows: The dynamics of globally minimizing orbits*, Bol. Soc. Bras. Mat, **28**, 141-153, (1997)

[23] Strang, G: *Maximal flow through a domain, Math. Programming* **26**, 123-143 (1983)

[24] Strang, G: *$L_1$ and $L_\infty$ approximations of vector fields in the plane*, in Lecture Notes in Num. Appl. Analysis **5** 273- 288 (1982)

[25] Rockafeller, R.T: *Convex Analysis*, Princeton, N.J, Princeton U. Press

[26] Villani C.: *Optimal Transport: Old and New*, Grundlehren der mathematischen Wissenschaften, (2008)

[27] Siburg, F: *The Principle of Least Action in geometry and Dynamics* , Lecture Notes in Mathematics **1844**, Springer,2004.

[28] Villani C.: *Optimal Transport: Old and New*, Grundlehren der mathematischen Wissenschaften, (2008)

[29] Wolansky, G: *Minimizers of Dirichlet functionals on the n-torus and the weak KAM theory*, Ann. Inst. H. Poincare' Anal. Non Line'aire **26**, 521–545. (2009)

[30] Wolansky G: *Extended least action principle for steady flows under a prescribed flux*, Calc. Var. Partial Differential Equations **31**, 277-296, (2008)