From Hopf-Lax formula to optimal weak transfer plan

Yan Shu[*]

May 28, 2019

Abstract

We study the properties of Hopf-Lax formula restricted to convex functions and provide a characterization of the optimal transfer plan for weak transport problems on the real line. On n dimensional real space, we also provide a sufficient condition on the potential function such that the optimal plan of the classical Monge-Kantorovich problem does not depend on the cost function. As a byproduct, we establish a link between the combinatorial object (the permutation polytope) and the Hamilton Jacobi equation.

key words: Weak transport costs; Hopf-Lax formula; Monge-Kantorovich problem; Hamilton Jacobi equation

1 Introduction

1.1 Monge-Kantorovich problem and weak transport cost

Throughout the paper, the space would be Euclidean space $\mathbb{R}$ or $\mathbb{R}^n$. We denote $\mathcal{P}(\mathbb{R}^n)$ the set of probability measures on $\mathbb{R}^n$, $\mathcal{P}_1(\mathbb{R}^n)$ the set of probability measures on $\mathbb{R}^n$ with finite first moment and $\|.\|$, the Euclidean norm.

Given two probability measures $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^n)$, recall that the classical Monge-Kantorovich problem is to minimize the following transportation cost:

$$\mathcal{T}_c(\mu, \nu) = \inf_\pi \int c(x, y) d\pi \in [0, \infty],$$

where $c : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$ is called the cost function and π is a measure coupling with marginal μ and ν. Readers can refer to [26] for an introduction of the optimal transport theory. One of the most interesting cases is that the cost function c is a convex function with respect to the distance d. One of the well-known results about the optimal transfer plan is due to Brenier [6] in 1991, he proved that on a finite dimensional real space, the optimal transfer plan is a gradient of some convex function. In particular, in one dimension, the optimal transfer plan does not depend on the cost function if the cost function is convex. We also refer readers to related results about the optimal transfer plan in [10], [18], [21].

---

[*]Modélisation aléatoire de Paris Ouest Nanterre La Défense(MODAL'X), Email: yan.shu.prof@gmail.com

Here we are interested in the optimal transfer plan of a variant of Monge-Kantorovich problem, which is introduced in [13], the general optimal transportation problem, defined by

$$\overline{\mathcal{T}}_c(\nu|\mu) = \inf_\pi \int c(x, p)\, \mu(dx). \tag{1.1}$$

The cost function c is defined on the space $\mathbb{R}^n \times \mathcal{P}(\mathbb{R}^n)$, the infimum runs over all couplings $\pi(dxdy) = p(x, dy)\mu(dx)$ of μ and ν and where $p(x, \cdot)$ denotes the disintegration kernel of π with respect to its first marginal. Further papers directly related to this problem include [2, 3, 8, 11, 12, 22, 23, 24].

In terms of random variables, one has the following interpretation:

$$\overline{\mathcal{T}}_c(\nu|\mu) = \inf \mathbb{E}\left(c(X, \mathbb{E}(Y|X))\right),$$

whereas

$$\mathcal{T}_c(\nu, \mu) = \inf \mathbb{E}\left(c(X, Y)\right),$$

where in both cases the infimum runs over all random vector (X, Y) such that X has law μ and Y has law ν. This general transport cost (1.1) plays an important role in the study of Talagrand type transport inequalities and some log-Sobolev inequalities, especially in a discrete space such as graphs or a subset of vector space. Those transport inequalities immediately yield concentration results and tensorization properties of the measure (see [8, 13, 12, 23, 24]).

Throughout the paper, the cost function is of form: $c(x, y) = \theta(x - y)$, where $\theta : \mathbb{R}^n \mapsto \mathbb{R}^+$ is a non-constant convex function vanishing at 0 and radially symmetric with respect to the origin. In one dimension, θ is even, and in n dimension, θ only depends on the distance between x and y. In what follows, taking two probability measures $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^n)$, the optimal weak transport cost of ν with respect to μ and cost function $\theta : \mathbb{R}^n \mapsto \mathbb{R}$ means

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \inf_\pi \int \theta\left(x - \int y\, p(x, dy)\right)\mu(dx)$$

where the infimum runs over all couplings $\pi(dxdy) = p(x, dy)\mu(dx)$ of μ and ν, and where $p(x, \cdot)$ denotes the disintegration kernel of π with respect to its first marginal. Since the θ is convex, by Jensen's inequality, one has

$$\overline{\mathcal{T}}_\theta(\nu|\mu) \leqslant \mathcal{T}_\theta(\nu, \mu).$$

This weak transport cost is deeply linked to the Monge-Kantorivich problem. Together with Gozlan, Roberto, Samson and Tetali, following Strassen's theorem [25], we proved in [12] that,

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \inf_{\nu_1 \preceq \nu} \mathcal{T}_\theta(\nu_1, \mu),$$

where $\preceq$ is the convex order, defined as:

$$\nu_1 \preceq \nu \qquad \Leftrightarrow \qquad \forall f\ convex, \int f d\nu_1 \leqslant \int f d\nu.$$

2

Moreover, in one dimension, the measure ν_1 which achieves the infimum does not depend on the choice of the convex cost function θ. As an application, one can deduce a completed characterization of a convex modified Log-Sobolev inequality [24].

1.2 Presentation of the results

It is natural to ask the following questions about the optimal transfer plan of weak transport problem:

- When the weak transport would be equal to the classical transport?

- Does the optimal coupling depend on the cost function θ?

We approach those questions by the following duality framework of the weak transport cost. Following [13, Theorem 2.11], for $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^n)$, and $\theta : \mathbb{R}^n \mapsto \mathbb{R}$ a convex function such that $\theta(x) \geqslant a\|x\| + b$ for some $a > 0$ and $b \in \mathbb{R}$, it holds

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \sup_f \int Q_1^\theta f d\mu - \int f d\nu \qquad (1.2)$$

where the supremum runs over all function f convex, Lipschitz and bounded from below, and where

$$Q_t^\theta f(x) = \inf_{y \in \mathbb{R}^n} \left\{ f(y) + t\theta\left(\frac{y-x}{t}\right) \right\}.$$

The above equation is often called the Hopf-Lax formula, in some references the operator Q_t^θ is also called the inf-convolution operator and often denoted simply by Q^θ for $t = 1$. The Hopf-Lax formula is known as the solution of a Hamilton Jacobi equation, and has been widely studied in many different contexts. We remark as well in the proof of [2, Theorem 5.5], Backhoff-Veraguas, Beiglbock and Pammer shown that a maximizer of equation (1.2) exists even with a weaker assumption.

In this paper, since the cost function θ is convex, real valued, positive and only depended on the euclidean norm, it satisfies the growth condition $\theta(x) \geqslant a\|x\| + b$ and the condition (A+) in [2]. Therefore we can apply the duality theorem and the existence of the maximizer here.

Apart from the introduction, this paper is divided into three sections. In section 2, the space will be the real line $\mathbb{R}$, and in section 3, the space will be $\mathbb{R}^n$ and we will present some applications in section 4.

In section 2, we stay in one dimension. We provide an equivalent condition for the equality between the weak transport cost and the classical Monge-Kantorovich transport cost (theorem 2.6), which states that

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\nu, \mu)$$

if and only if the difference between the inverse cumulative functions of μ and ν is non-decreasing, precisely, the function

$$F_\mu^{-1} - F_\nu^{-1}$$

is non-decreasing. This equivalence also has been obtained recently by Gozlan and Juillet in a slightly different form in [11] and by Backhoff Veraguas, Beiglboeck and Pammer in [3].

Furthermore, according to Brenier [6], in one dimension, the optimal mapping of a classical transport does not depend on the cost function as soon as the cost function is convex. In [12], the same result is obtained for weak transport cost. We will give a new proof of this result in section 2, using a very different argument. During this approach, a byproduct about the Hamilton-Jacobi equation is obtained, which might have its own interest (theorem 2.11).

In section 3, the space will be $\mathbb{R}^n$. We will show in theorem 3.4 an extension of the results of section 2, which provide equivalence conditions (under some smoothness properties) such that $\overline{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\mu, \nu)$ holds. We will define Property $\mathcal{F}$, which reduces to convexity in one dimension and plays the role of the convexity in n dimensions. The condition that there is equality between the weak transport cost and the classical transport cost is deeply related to Property $\mathcal{F}$.

In section 4, we briefly explain some applications on the infimum convolution inequality introduced by Maurey in [17].

2 Weak transport on the line

In this section we will focus on the real line. A cost function in this section is a function $\theta : \mathbb{R} \mapsto \mathbb{R}^+$, convex, even, satisfying $\theta(0) = 0$.

We remind here the notion of cyclical monotone of a multivalued mapping [20, Page 238]. A multivalued mapping ρ from $\mathbb{R}^n$ to $\mathbb{R}^n$ is called cyclically monotone if one has

$$\langle x_1 - x_0, x_0^* \rangle + \langle x_2 - x_1, x_1^* \rangle + \cdots + \langle x_0 - x_m, x_m^* \rangle \leqslant 0.$$

for any set of pairs (x_i, x_i^*), $i = 0, 1, ..., m$ (m arbitrary) such that $x_i^* \in \rho(x_i)$. The property of cyclical monotone characterizes the gradient of a convex function according to [20, Theorem 5.6], which states that, ρ is cyclically monotone if and only if there exist a convex proper function f on $\mathbb{R}^n$ such that $\rho \subset \partial f$.

2.1 A remark on Hopf-Lax formula

We begin with some development of the Hopf-Lax formula. The key observation is the following lemma.

Lemma 2.1. *Let I, J be two subset of $\mathbb{R}$ with strictly positive Lebesgue measure. Given $t > 0$, let $T_t : I \subset \mathbb{R} \to J \subset \mathbb{R}$ be a real valued function. Then the following statements hold.*

(i) If there exists a real-valued convex function f defined on $\mathbb{R}$ and a strictly convex cost θ (recall that a cost function here should be positive and vanishing at 0) such that for all $x \in I$, it holds

$$Q_t^\theta f(x) = f(T_t(x)) + t\theta\left(\frac{T_t(x) - x}{t}\right). \tag{2.2}$$

4

then T_t is non-decreasing and $x \mapsto T_t(x) - x$ is non-increasing.

(ii) Inversely, taking a function T_t defined on $I \subset \mathbb{R}$, if T_t is non-decreasing and $x \in J \mapsto T_t(x) - x$ is non-increasing, then for all convex cost θ, there exists a closed proper convex function f, such that (2.2) holds for all $x \in I$.

Proof. (i). Given $t > 0$ and $x \in I$, the function $G_x : u \mapsto f(u) + t\theta \left(\frac{u-x}{t} \right)$ is defined on $\mathbb{R}$. According to convexity of f and θ, G is strictly convex. Therefore, it has at most one minimum. Since a convex function always has a right derivative and a left derivative, then it holds for all $x \in I$:

$$0 \in [\partial_- G_x(T_t(x)), \partial_+ G_x(T_t(x))]. \tag{2.3}$$

We deduce that for all $x \in I$:

$$\partial_+ f(T_t(x)) + \partial_+ \theta \left(\frac{T_t(x) - x}{t} \right) \geq 0.$$

We will first prove that T_t is non-decreasing. For any $x, y \in I$, $x < y$, since θ is strictly convex, we deduce that

$$\partial_+ G_x(T_t(y)) = \partial_+ f(T_t(y)) + \partial_+ \theta \left(\frac{T_t(y) - x}{t} \right)$$

$$> \partial_+ f(T_t(y)) + \partial_+ \theta \left(\frac{T_t(y) - y}{t} \right) \geq 0 \geq \partial_+ G_x(T_t(x)).$$

Together with the strict convexity of G_x, we deduce that $T_t(y) \geq T_t(x)$.

Now we turn to prove that $x \mapsto T_t(x) - x$ is non-increasing on I. We will show that for $x, y \in I$ such that $x < y$, it holds $T_t(x) - x \geq T_t(y) - y$. Since T_t is non-decreasing, $T_t(x) \leq T_t(y)$.

If $T_t(x) = T_t(y)$, it follows immediately $T_t(x) - x > T_t(y) - y$.

Now we assume that $T_t(x) < T_t(y)$. Since f is convex, it holds

$$\partial_+ f(T_t(x)) \leq \partial_- f(T_t(y)). \tag{2.4}$$

Now applying (2.3) for x and y, we deduce:

$$\partial_+ G_x(T_t(x)) \geq 0 \geq \partial_- G_y(T_t(y)). \tag{2.5}$$

Combining equation (2.4) and (2.5), we deduce

$$\partial_+ \theta \left(\frac{T_t(x) - x}{t} \right) \geq \partial_- \theta \left(\frac{T_t(y) - y}{t} \right).$$

By convexity of θ, it holds $T_t(x) - x \geq T_t(y) - y$. Thus $x \mapsto T_t(x) - x$ is non-increasing on I.

(ii). Assume that the function T_t is non-decreasing on I and the function $x \mapsto T_t(x) - x$ is non-increasing on I. Define a multi valued mapping ρ from $\mathbb{R}$ to $\mathbb{R}$ such that

$$y^* \in \rho(y) \Leftrightarrow y^* = -\partial_+ \theta \left(\frac{y - x}{t} \right),$$

for some x such that $T_t(x) = y$.

For any set of pairs (y_i, y_i^*), $i = 0, 1, ..., m$, (m arbitrary) such that $y_i^* \in \rho(y)$, there exist $x_0, ..., x_m$, such that $T_t(x_i) = y_i$. We claim that $y_i < y_j$ implies $y_i^* \leqslant y_j^*$, for any $i, j \in \{0, 1, ..., m\}$. In fact, if $y_i < y_j$, then by monotony of T_t, it holds $x_i < x_j$. Since $x \mapsto T_t(x) - x$ is non-increasing, we deduce that $T_t(x_i) - x_i \geqslant T_t(x_j) - x_j$. Therefore, by convexity of θ,

$$y_i^* = -\partial_+ \theta \left(\frac{T_t(x_i) - x_i}{t} \right) \leqslant -\partial_+ \theta \left(\frac{T_t(x_j) - x_j}{t} \right) = y_j^*.$$

Therefore, according to rearrangement inequality, it holds:

$$\langle y_1 - y_0, y_0^* \rangle + \langle y_2 - y_1, y_1^* \rangle + \cdots + \langle y_0 - y_m, y_m^* \rangle$$
$$= \sum_{i=0}^{m} y_i y_i^* - \sum_{i=0}^{m} y_i y_{i+1} \leqslant 0,$$

where we denote $y_{m+1} = y_0$ and in dimension one, the bracket $\langle, \rangle$ is the multiplication. Thus, ρ is cyclically monotone. According to [20, Page 238, Theorem 24.8], there exists a closed proper convex function f (thus lower semi-continuous) defined on $\mathbb{R}$, such that for all $y \in \mathbb{R}$, $\rho(y) \subset \partial f(y)$ ($\rho(y)$ could be empty). Therefore, for all $x \in I$, it holds

$$-\partial_+ \theta \left(\frac{T_t(x) - x}{t} \right) \in [\partial_- f(T_t(x)), \partial_+ f(T_t(x))],$$

which leads to

$$0 \in \left[\partial_- f(T_t(x)) + \partial_- \theta \left(\frac{T_t(x) - x}{t} \right), \partial_+ f(T_t(x)) + \partial_+ \theta \left(\frac{T_t(x) - x}{t} \right) \right].$$

Thus equation (2.2) holds. The conclusion follows. $\qquad\square$

2.2 Characterization of equality between weak transport cost and transport cost

We firstly recall the definition of cumulative distribution function and its inverse.

For a probability measure $\mu \in \mathcal{P}(\mathbb{R})$, denote F_μ the cumulative distribution function of μ, i.e.

$$F_\mu(x) = \mu(-\infty, x],$$

and define the generalized inverse of F_μ^{-1} by

$$F_\mu^{-1}(t) = \inf\{x \in \mathbb{R}; F_\mu(x) > t\}.$$

Theorem 2.6. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$ and let the cost function $\theta : \mathbb{R} \mapsto \mathbb{R}^+$ be even, strictly convex and vanishing at 0. Assume that μ and ν are absolutely continuous with respect to Lebesgue measure. Assume that the weak transport $\overline{\mathcal{T}}_\theta(\nu|\mu)$ is finite, then*

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\mu, \nu)$$

if and only if $F_\mu^{-1} - F_\nu^{-1}$ is non-decreasing.

In one dimension, given two probability measures μ and ν which are absolutely continuous with respect to Lebesgue measure, the optimal transport mapping for transport problem is transporting mass from location $F_\mu^{-1}(x)$ to location $F_\nu^{-1}(x)$. This mapping is in fact the monotone rearrangement T from μ to ν. We will play with the Kantorovich duality formula related to this mapping T for the weak transport problem.

Proof. We assume at first that $F_\mu^{-1} - F_\nu^{-1}$ is non-decreasing. Denote T the monotone rearrangement from μ to ν and denote $I := \text{supp}(\mu)$ the support of μ and $J := \text{supp}(\nu)$ the support of ν. Thus T is an one-to-one mapping from I to J, and for all $x \in I$,

$$F_\mu(x) = \mu(-\infty, x) = \nu(-\infty, T(x)) = F_\nu(T(x)). \tag{2.7}$$

Since μ, ν are absolutely continuous, F_μ^{-1} and F_μ^{-1} are well defined on $(0,1)$. Now given $u \in (0,1)$, denote $u = F_\mu(x) = F_\nu(T(x))$. According to the equality (2.7), it holds

$$F_\nu^{-1}(u) - F_\mu^{-1}(u) = T(x) - x.$$

According to our hypothesis, $x \mapsto T(x) - x$ is non-increasing. Moreover, notice that T is the monotone rearrangement from μ to ν, T is non-decreasing. We can apply lemma 2.1, and there exists a closed proper convex function f (thus lower semi-continuous), such that for all $x \in I$:

$$Q^\theta f(x) = \inf_{y \in \mathbb{R}} \{f(y) + \theta(y - x)\} = f(T(x)) + \theta(T(x) - x).$$

Since f is proper, there exist $x_0 \in \mathbb{R}$ such that $f(x_0) < \infty$. By definition of Qf, it holds

$$Q^\theta f(x) = f(T(x)) + \theta(T(x) - x) \leqslant f(x_0) + \theta(x_0 - x).$$

Thus, for $R > |x_0|$ and for all $x \in I \cap [-R, R]$,

$$f(T(x)) \leqslant f(x_0) + \theta(x_0 - x) - \theta(T(x) - x) \leqslant f(x_0) + \theta(2R).$$

Therefore f is a lower semi-continuous, bounded and lipschitz on the interior of $I \cap [-R, R]$.

By Kantorovich duality from [13], we have

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \sup_{\varphi \ convex, Lip, \ l.s.c.} \int Q^\theta \varphi d\mu - \int \varphi d\nu$$

$$\geqslant \int_{I \cap [-R,R]} Q^\theta f d\mu - \int_{I \cap [-R,R]} f d\nu = \int_{I \cap [-R,R]} \theta(T(x) - x) d\mu$$

$$= \mathcal{T}_\theta(\mu, \nu) - \varepsilon_R \geqslant \overline{\mathcal{T}}_\theta(\nu|\mu) - \varepsilon_R.$$

Now let $R \to \infty$, $\varepsilon_R \to 0$, and we obtain

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\mu, \nu).$$

Now assume that $\overline{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\mu,\nu)$. According to [2, Theorem 5.5], for $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$, if $\overline{\mathcal{T}}_\theta(\nu|\mu)$ is finite, there exists a convex function f such that

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \int Q^\theta f d\mu - \int f d\nu.$$

Thus,

$$\begin{aligned}
\overline{\mathcal{T}}_\theta(\nu|\mu) &= \int Q^\theta f d\mu - \int f d\nu \\
&\leqslant \int f(T(x)) + \theta(T(x) - x) d\mu - \int f d\nu = \int \theta(T(x) - x) d\mu \\
&= \mathcal{T}_\theta(\mu,\nu).
\end{aligned}$$

Therefore, $Q^\theta f(x) = f(T(x)) + \theta(T(x) - x)$ μ almost-surely. Since f is convex and θ is strictly convex, according to lemma 2.1 (i), we obtain that $T(x) - x$ is non-increasing μ almost surely, together with the monotonicity of the function F_μ^{-1}, the conclusion follows. $\qquad\square$

2.3 Weak transfer plan

In this section we give an alternative proof of a theorem in [12], as a consequence of lemma 2.1, which is the following:

Theorem 2.8. [12, Theorem 1.4] Let α, β and θ be convex cost functions satisfying $\alpha + \beta = \theta$. Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$. Assume that $\overline{\mathcal{T}}_\theta(\nu|\mu)$ is finite, it holds

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \overline{\mathcal{T}}_\alpha(\nu|\mu) + \overline{\mathcal{T}}_\beta(\nu|\mu). \tag{2.9}$$

We need to prove the following proposition at first.

Proposition 2.10. Let $\alpha, \beta, \theta : \mathbb{R} \to \mathbb{R}$ be convex cost functions, even and of class C^1 satisfying $\alpha + \beta = \theta$. We assume that θ is strictly convex. Then for all convex function $f : \mathbb{R} \to \mathbb{R}$ bounded from below of class C^1, there exists convex functions $f_1, f_2 : \mathbb{R} \mapsto \mathbb{R}$, bounded from below and of class C^1, such that

$$f = f_1 + f_2$$

and for all $t > 0$

$$Q_t^\theta f = Q_t^\alpha f_1 + Q_t^\beta f_2$$

more precisely,

$$f_1(x) = a + \int_0^x \alpha' \circ \theta'^{-1} \circ f'(u) du,$$

with any constant $a \in \mathbb{R}$.

From this proposition, combining with the fact that $Q_t^\theta f$ is in fact the solution of the Hamilton Jacobi equation (2.12) [7], one deduces immediately of the following theorem.

Theorem 2.11. *Let $\theta : \mathbb{R} \to \mathbb{R}$ be a strictly convex cost function with super-linear growth of class C^1 (i.e. $\theta'(x)$ goes to ∞ as x goes to ∞) and θ^* be the Legendre transform of θ. Consider the following Hamilton Jacobi equation:*

$$\begin{cases} \partial_t v(x,t) + \theta^*(\partial_x v(x,t)) = 0 & (x,t) \in \mathbb{R} \times (0, \infty) \\ v(x,0) = f(x) & x \in \mathbb{R}. \end{cases} \tag{2.12}$$

Assume that the initial function $f : \mathbb{R} \to \mathbb{R}$ is convex bounded from below of class C^1, then for all convex cost function θ_1, θ_2, with super-linear growth and satisfying $\theta_1 + \theta_2 = \theta$, there exists v_1, v_2 and f_1, f_2, such that for $i = 1, 2$, it holds

$$\begin{cases} \partial_t v_i(x,t) + \theta_i^*(\partial_x v_i(x,t)) = 0 & (x,t) \in \mathbb{R} \times (0, \infty) \\ v_i(x,0) = f_i(x) & x \in \mathbb{R}, \end{cases}$$

$v = v_1 + v_2$ *and* $f = f_1 + f_2$.

We begin with a lemma.

Lemma 2.13. *Assume that θ is strictly convex, even and of class C^1. Let f be a convex function bounded from below and of class C^1. For $t > 0$, denote $I := \{x \in \mathbb{R} \mid f'(x) \in \theta'(\mathbb{R})\}$. Define the mapping U_t as*

$$U_t : x \in I \mapsto x + t\theta'^{-1} \circ f'(x) \in \mathbb{R}.$$

Then $T_t := U_t^{-1}$ satisfies (2.2).

Proof. Since θ is even, strictly convex and of class C^1, $\theta'(\mathbb{R})$ contains a neighborhood of 0, combining with the fact that f is bounded from below, convex and C^1, we deduce that I is not an empty set. Since θ is strictly convex, $\theta' : \mathbb{R} \to \theta'(\mathbb{R})$ is a bijection and it is strictly increasing. Thus $\theta'^{-1} : \theta'(\mathbb{R}) \to \mathbb{R}$ is well defined as well as U_t. It is easy to see that U_t is strictly increasing and continuous. Now we will show that the image set $U_t(I) = \mathbb{R}$ and U_t is in fact a bijection from I to $\mathbb{R}$.

Assume first that $\inf I = -\infty$, then it is easy to see that $\lim_{x \to -\infty} U_t(x) = -\infty$. Now assume that $\inf I > -\infty$. Then for $x \searrow \inf I$, $f'(x) \searrow \theta(-\infty)$, thus $\theta'^{-1} f'(x) \to -\infty$. The same argument holds for $\sup I$. Therefore, $U_t : I \to \mathbb{R}$ is bijective and strictly increasing. Thus $U_t^{-1} : \mathbb{R} \to I$ is well defined and strictly increasing.

It remains to show that $U_t^{-1}(x)$ is the point achieving the infimum for $Q_t^\theta f(x)$. For all $x \in \mathbb{R}$, let $y = U_t^{-1}(x)$, using the fact that θ is even, it holds

$$\theta'\left(\frac{y-x}{t}\right) = \theta'\left(\frac{y - U_t(y)}{t}\right) = -\theta'\left(\theta'^{-1} \circ f'(y)\right) = -f'(y).$$

According to (2.3), it holds $y = T_t(x)$. $\qquad \square$

Now we are in position to prove proposition 2.10.

9

Proof of proposition 2.10. We define f_1 as following: for all $y \in I$,

$$f_1(y) := \int_0^y \alpha' \circ \theta'^{-1} \circ f'(u) du.$$

and (in the case that $I \neq \mathbb{R}$) f_1 is affine when $y > \sup I$, and $y < \inf I$ with $f_1' = f_1'(\sup I) := \lim_{x \to \sup I} f_1'(x)$ and $f_1' = f_1'(\inf I)$ respectively.

We observe that f_1 is convex: for all $x \in I$, $f_1' = \alpha' \circ \theta'^{-1} \circ f'$ is non-decreasing, and for $x \in \mathbb{R} \setminus I$, f_1' is constant which equals to $f_1' = f_1'(\sup I)$ and $f_1' = f_1'(\inf I)$ respectively. It follows that f_1' is non-decreasing on $\mathbb{R}$.

Given $t > 0$ and $x \in \mathbb{R}$, denote $y = T_t(x) \in I$ and it holds

$$f_1'(y) = \alpha' \circ \theta'^{-1} \circ f'(y)$$
$$= -\alpha' \circ \theta'^{-1} \circ \theta' \left(\frac{y-x}{t} \right) = -\alpha' \left(\frac{y-x}{t} \right).$$

We deduce from equation (2.3) that

$$Q_t^\alpha f_1(x) = f_1(T_t(x)) + t\alpha \left(\frac{T_t(x) - x}{t} \right). \tag{2.14}$$

Now let $f_2 = f - f_1$, together with $\theta = \alpha + \beta$, we deduce

$$f_2'(x) = \begin{cases} f' - f_1'(\inf I) & \forall x \in (-\infty, \inf I] \\ \beta' \circ \theta'^{-1} \circ f'(x) & \forall x \in I \\ f' - f_1'(\sup I) & \forall x \in [\sup I, \infty) \end{cases},$$

which is non-decreasing. Thus f_2 is convex.

On the other hand, since $f_1'(y) = -\alpha' \left(\frac{y-x}{t} \right)$ and $f'(y) = -\theta' \left(\frac{y-x}{t} \right)$, it holds that

$$f_2'(y) = -\beta' \left(\frac{y-x}{t} \right).$$

According to equation (2.3) and we have

$$Q_t^\beta f_2(x) = f_2(T_t(x)) + t\beta \left(\frac{T_t(x) - x}{t} \right). \tag{2.15}$$

It immediately yields

$$Q_t^\theta f = Q_t^\alpha f_1 + Q_t^\beta f_2.$$

$\square$

Now we are in position to proof theorem 2.8.

Proof of theorem 2.8. We first prove the case that the convex cost functions α,β and θ are C^1:

We assume at first that $\theta = \alpha + \beta$ is strictly convex.

Observe that by the definition of $\overline{\mathcal{T}}$, it is easy to see that

$$\overline{\mathcal{T}}_\theta(\nu|\mu) \geqslant \overline{\mathcal{T}}_\alpha(\nu|\mu) + \overline{\mathcal{T}}_\beta(\nu|\mu). \tag{2.16}$$

Now we turn to prove the inverse inequality.

According to Proposition 2.10, for all convex function f bounded from below and of class C^1, there exist f_1 and f_2 such that it holds:

$$\int Q^\theta f d\mu - \int f d\nu = \int Q^\alpha f_1 d\mu - \int f_1 d\nu + \int Q^\beta f_2 d\mu - \int f_2 d\nu$$
$$\leqslant \overline{\mathcal{T}}_\alpha(\nu|\mu) + \overline{\mathcal{T}}_\beta(\nu|\mu).$$

We take the supremum over all convex function f bounded from below and of class C^1 and by the duality formula (1.2), we get

$$\overline{\mathcal{T}}_\theta(\nu|\mu) \leqslant \overline{\mathcal{T}}_\alpha(\nu|\mu) + \overline{\mathcal{T}}_\beta(\nu|\mu).$$

The conclusion follows with the inverse inequality (2.16).

Now assume that θ is not strictly convex. Since $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$, there exists a strictly convex cost function γ, such that $\int \gamma(x-c)\mu(dx)$ and $\int \gamma(x-c)\nu(dx)$ are finite for all $c \in \mathbb{R}$. Therefore $\overline{\mathcal{T}}_{\theta+\gamma}(\nu|\mu)$ is finite. Since $\alpha + \gamma$, $(\alpha+\gamma)+\beta$ are strictly convex, it holds:

$$\overline{\mathcal{T}}_{\theta+\gamma}(\nu|\mu) = \overline{\mathcal{T}}_\theta(\nu|\mu) + \overline{\mathcal{T}}_\gamma(\nu|\mu). \tag{2.17}$$

and

$$\overline{\mathcal{T}}_{\theta+\gamma}(\nu|\mu) = \overline{\mathcal{T}}_{(\alpha+\gamma)+\beta}(\nu|\mu) = \overline{\mathcal{T}}_{\alpha+\gamma}(\nu|\mu) + \overline{\mathcal{T}}_\beta(\nu|\mu)$$
$$= \overline{\mathcal{T}}_\alpha(\nu|\mu) + \overline{\mathcal{T}}_\beta(\nu|\mu) + \overline{\mathcal{T}}_\gamma(\nu|\mu). \tag{2.18}$$

Combining equation (2.17) and (2.18), we deduce that the equality (2.9) holds for the case α, β and θ are C^1.

Now for general convex cost α, β and θ, for all $\varepsilon > 0$ there exist cost functions $\alpha_\varepsilon, \beta_\varepsilon$ and θ_ε satisfying $\theta_\varepsilon = \alpha_\varepsilon + \beta_\varepsilon$ of class C^1 such that

$$\|\alpha_\varepsilon - \alpha\|_\infty + \|\beta_\varepsilon - \beta\|_\infty + \|\theta_\varepsilon - \theta\|_\infty < \varepsilon.$$

We deduce that

$$\overline{\mathcal{T}}_\alpha(\nu|\mu) + \overline{\mathcal{T}}_\beta(\nu|\mu) \leqslant \overline{\mathcal{T}}_\theta(\nu|\mu)$$
$$\leqslant \overline{\mathcal{T}}_{\theta_\varepsilon}(\nu|\mu) + \varepsilon = \overline{\mathcal{T}}_{\alpha_\varepsilon}(\nu|\mu) + \overline{\mathcal{T}}_{\beta_\varepsilon}(\nu|\mu) + \varepsilon$$
$$\leqslant \overline{\mathcal{T}}_\alpha(\nu|\mu) + \overline{\mathcal{T}}_\beta(\nu|\mu) + 2\varepsilon.$$

Let ε goes to 0 and the theorem follows.

$\square$

2.4 An alternative approach

The proof of theorem 2.6 in [12] is based on properties of convex ordering and Rado's theorem. In this section, we provide another way to understand theorem 2.6 in aspect of [12].

Here we only recall some necessary definitions and properties of the convex ordering and majorization of vectors. We refer the interested reader to [16], [14] and [19] for further results and bibliographic references related to properties of inequalities, and to [12, 11, 3] for recent developments related to the optimal weak transfer plan.

Definition 2.19 (Convex order). *Given $\nu_1, \nu_2 \in \mathcal{P}_1(\mathbb{R})$, we say that ν_2 dominates ν_1 in the convex order, and write $\nu_1 \preceq \nu_2$, if for all convex functions f on $\mathbb{R}$, $\int_{\mathbb{R}} f \, d\nu_1 \leqslant \int_{\mathbb{R}} f \, d\nu_2$.*

Definition 2.20 (Majorization of vectors). *Let $a, b \in \mathbb{R}^n$, one says that a is majorized by b if the sum of the largest j components of a is less than or equal to the corresponding sum of b, for every j, and if the total sum of the components of both vectors are equal.*

Assuming that the components of $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ are in non-decreasing order (*i.e.* $a_1 \leqslant a_2 \leqslant \ldots \leqslant a_n$ and $b_1 \leqslant b_2 \leqslant \ldots \leqslant b_n$), a is majorized by b, if

$$a_n + a_{n-1} + \cdots + a_{n-j+1} \leq b_n + b_{n-1} + \cdots + b_{n-j+1}, \qquad \text{for } j = 1, \ldots, n-1,$$

and $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i$.

The next proposition recalls the link between majorization of vectors and convex ordering.

Proposition 2.21. *[12] Let $a, b \in \mathbb{R}^n$ and set $\nu_1 = \frac{1}{n} \sum_{i=1}^n \delta_{a_i}$ and $\nu_2 = \frac{1}{n} \sum_{i=1}^n \delta_{b_i}$. The following statements are equivalent.*

(i) a is majorized by b.

(ii) ν_1 is dominated by ν_2 for the convex order. In other words, for every convex function $f : \mathbb{R} \to \mathbb{R}$, it holds that $\sum_{i=1}^n f(a_i) \leq \sum_{i=1}^n f(b_i)$.

Thanks to the above proposition and with a slight abuse of notation, we will also write $a \preceq b$ when a is majorized by b.

Now we are able to prove an alternative version of theorem 2.6. We shall focus on measure μ_n of form $\frac{1}{n} \sum_{i=1}^n \delta_{x_i}$. For general measure, one can consider it as a limit of μ_n as n goes to ∞, readers can refer to [12] for rigorous justification. Here is a discrete version of theorem 2.6 is telling the following:

Theorem 2.22 (Discrete version of theorem 2.6). *Let $\mu = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ and $\nu = \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$, where x_i and y_i are in non-decreasing order. Assume that θ is a strictly convex cost function. The following statements are equivalent.*

(i) The function $i \mapsto x_i - y_i$ is non-decreasing.

(ii) $\overline{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\nu, \mu)$.

Proof. Observe that the optimal transfer plan of $\mathcal{T}_\theta(\nu, \mu)$ sends x_i to y_i since x_i and y_i are in non-decreasing order. As a consequence,

$$\mathcal{T}_\theta(\nu, \mu) = \frac{1}{n} \sum_{i=1}^{n} \theta(x_i - y_i).$$

On the other hand, denoting $x = (x_1, ..., x_n), y = (y_1, ..., y_n) \in \mathbb{R}^n$, according to [12], the following holds:

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \inf_{\nu_1 \preceq \nu} \mathcal{T}_\theta(\nu_1, \mu) = \inf_{y' \preceq y} \frac{1}{n} \sum_{i=1}^{n} \theta(x_i - y'_i).$$

Thus, item (ii) is equivalent to

$$\sum_{i=1}^{n} \theta(x_i - y_i) = \inf_{y' \preceq y} \sum_{i=1}^{n} \theta(x_i - y'_i). \tag{2.23}$$

Now it is enough to prove that (2.23) is equivalent to (i).

(i) $\Rightarrow$ (ii): For any $y' \preceq y$, it holds

$$\sum_{i=k}^{n} y'_i \leqslant \sum_{i=k}^{n} y_i \qquad \forall 1 \leqslant k \leqslant n.$$

It follows that

$$\sum_{i=k}^{n} x_i - y_i \leqslant \sum_{i=k}^{n} x_i - y'_i \qquad \forall 1 \leqslant k \leqslant n.$$

Thus $x - y \preceq x - y'$ for all $y' \preceq y$, which leads to (2.23).

(ii) $\Rightarrow$ (i): Assume that (ii) holds and (i) does not. Let i be the smallest integer such that $x_i - y_i > x_{i+1} - y_{i+1}$. Thus $y_{i+1} - y_i - (x_{i+1} - x_i) = 2\varepsilon > 0$.

Now define $y' \in \mathbb{R}^n$ with $y'_j = y_j$ for $j \neq i, i + 1$ $y'_i = y_i + \varepsilon, y'_{i+1} = y_{i+1} - \varepsilon$. It is easy to see that $y' \preceq y$ and by strict convexity of θ:

$$\sum_{i=1}^{n} \theta(x_i - y_i) - \sum_{i=1}^{n} \theta(x_i - y'_i)$$

$$= \theta(x_i - y_i) + \theta(x_{i+1} - y_{i+1}) - \theta(x_i - y_i - \varepsilon) - \theta(x_{i+1} - y_{i+1} + \varepsilon)$$

$$= \theta(x_i - y_i) + \theta(x_{i+1} - y_{i+1}) - 2\theta \left(\frac{(x_i - y_i) + (x_{i+1} - y_{i+1})}{2} \right)$$

$$> 0$$

which is a contradiction to (2.23). $\qquad \square$

3 Weak transport in n dimensions

This section is devoted to extending theorem 2.11 and theorem 2.6 of the previous section in n dimensions for $n \geqslant 2$. In this section, we will focus on the space $\mathbb{R}^n$ and an open subset $\Omega \subset \mathbb{R}^n$, which correspond to the support of an absolutely continuous measure w.r.t Lebesgue measure. We assume that all cost functions are convex, only depending on Euclidean distance $\|.\| := \|.\|_2$ and vanishing in 0. We denote $\mathcal{L}$ for the Lebesgue measure.

In one dimension, convexity of functions plays a central role in the proofs, but in n dimensions, we will replace the convexity by another stronger property, we call it property $\mathcal{F}$, which is defined as following.

Definition 3.1. *Let $\Omega \subset \mathbb{R}^n$ be a convex open set satisfying $\mathcal{L}(\Omega) > 0$. We say that $f : \Omega \mapsto \mathbb{R}$ satisfies the Property $\mathcal{F}$ on Ω if the following conditions hold:*

  1. f is convex on Ω.

  2. For almost all $x \in \Omega$, there exists λ, such that $\mathrm{Hess}f(x)\nabla f(x) = \lambda \nabla f(x)$.

Remark 3.2. *Theorem of Alexandrov [1] guarantees that the Hessian of a convex function f is well defined on $\mathbb{R}^n$ almost everywhere.*

Remark 3.3. *For dimension $n = 1$, the Property $\mathcal{F}$ is the convexity.*

Now let $f : \Omega \to \mathbb{R}$ be a continuous function. Denote

$$\mathcal{H}_f := \{x \in \Omega, \mathrm{Hess}f(x) \ exists\}.$$

Due to regularity issues, we introduce the following smoothness assumption **(SA)**. **(SA)**: The interior of $\mathcal{H}_f$, denoted by $\mathcal{H}_f{}^\circ$, is simply connected, and satisfies

$$\mathcal{L}(\Omega \setminus \mathcal{H}_f{}^\circ) = 0.$$

Now we are ready to state the main theorem of this section, which is an extension of theorem 2.6 in n dimensions.

Theorem 3.4. *Let $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^n)$ be two probability measures absolutely continuous with respect to Lebesgue measure. Assume that the support of the measure ν is a convex set Ω. Denote g the convex function such that ∇g is the transfer plan for Monge-Kantorovich problem of quadratic cost, and satisfying $\nabla g \# \nu = \mu$. If g satisfies the smoothness assumption* (SA), *then the following assertions hold.*

  (i) If there exists a two times differentiable cost function θ such that $x\theta''(u) - \theta'(u) \neq 0$ for all $u > 0$ and

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\mu, \nu) = \int \theta(\|x - \nabla g(x)\|)d\nu < \infty.$$

  Then the function $x \mapsto g(x) - \frac{1}{2}\langle x, x \rangle$ satisfies the Property $\mathcal{F}$ on Ω, and ∇g is an optimal transfer plan of $\mathcal{T}_\alpha$ for all convex cost α (such that $\overline{\mathcal{T}}_\theta(\nu|\mu)$ is finite). Since the transport mapping ∇g of the quadratic cost is unique, ∇g is the unique optimal transport plan in common for all convex cost α.

(ii) Inversely, if $x \mapsto g(x) - \frac{1}{2}\langle x, x \rangle$ is strictly convex and satisfies Property $\mathcal{F}$, then for all convex cost θ such that $\overline{\mathcal{T}}_\theta(\nu|\mu)$ is finite, ∇g is an optimal transfer plan of $\mathcal{T}_\theta$ for all convex cost θ and it holds

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\mu, \nu).$$

Remark 3.5. *The existence and uniqueness of the optimal transport mapping ∇g of the quadratic cost is guaranteed by the Benamou-Brenier theorem [4].*

Remark 3.6. *In one dimension, adapting the notation in theorem 2.6, the mapping ∇g is correspond to $F_\mu \circ F_\nu^{-1}$. A simple computation shows that the convexity of $g(x) - \frac{1}{2}x^2$ is equivalent to the fact that $F_\mu^{-1} - F_\nu^{-1}$ is non-decreasing. Therefore the equality between the weak transport cost and the Monge-Kantorovich transport cost is equivalence to the convexity of $g(x) - \frac{1}{2}x^2$. The latter theorem shows that with the additional smoothness assumption* (SA)*, this equivalent still holds. Heuristically, I do believe that it holds true without* (SA)*, but the regularity issue is very delicate and there is still a lot to investigate on.*

In order to prove the theorem, we begin with studying Property $\mathcal{F}$.

3.1 Characterization of Property $\mathcal{F}$

Let us begin with some properties of the set $\mathcal{F}$:

Lemma 3.7. *Let $\Omega \subset \mathbb{R}^n$ be a convex open set. Let $f : \Omega \to \mathbb{R}$ be a continuous function. Denote $u = \|\nabla f\|$. Then for all $x \in \mathcal{H}_f$, the following conditions are equivalent:*

(i) there exists $\lambda \in \mathbb{R}$ such that

$$\mathrm{Hess} f(x) \nabla f(x) = \lambda \nabla f(x).$$

(ii) for all $1 \leqslant i \leqslant j \leqslant n$, if $\|\nabla f\|(x) \neq 0$, it holds

$$\partial_i u \partial_j f = \partial_j u \partial_i f.$$

From this lemma, we can construct some examples of f such that $f \in \mathcal{F}$, by simply checking (ii).

Example 3.8. *In n dimensions, following functions satisfy Property $\mathcal{F}$ on $\mathbb{R}^n$:*

1. *Linear forms.*

2. *Functions of form $x \in \mathbb{R}^n \mapsto g(\|x\|)$ where $g : \mathbb{R} \to \mathbb{R}$ is a convex function with $g(0) = 0$.*

3. *Functions of form $x \in \mathbb{R}^n \mapsto a\|x\|^2 + L(x)$ with $a > 0$ and $L : \mathbb{R}^n \to \mathbb{R}$ being a linear form.*

Proof of lemma 3.7. $(i) \Rightarrow (ii)$:

u is well defined on $\mathcal{H}_f$. Suppose that (i) holds. For all $x \in \mathcal{H}_f$ with $\nabla f(x) \neq 0$, it holds for all $1 \leqslant i \leqslant j \leqslant n$,

$$\partial_i u = \frac{1}{u} \sum_{k=1}^n \partial_k f \partial_{ki} f = \frac{\lambda}{u} \partial_i f. \tag{3.9}$$

Thus

$$\partial_i u \partial_j f = \frac{\lambda}{u} \partial_i f \partial_j f = \partial_j u \partial_i f.$$

$(ii) \Rightarrow (i)$: For all $x \in \mathcal{H}_f$ such that $\nabla f(x) \neq 0$, there exists j such that $\partial_j f \neq 0$. Let $\lambda = \partial_j u / \partial_j f$, (ii) implies that for all $1 \leqslant i \leqslant n$,

$$\partial_i u = \lambda \partial_i f. \tag{3.10}$$

Computing the differential of u, it holds

$$\partial_i u = \frac{1}{u} \sum_{k=1}^n \partial_k f \partial_{ki} f.$$

Together with (3.10), it holds for all $1 \leqslant i \leqslant n$

$$\sum_{k=1}^n \partial_{ki} f \partial_k f = \lambda u \partial_i f,$$

which means exactly $\mathrm{Hess} f \, \nabla f = \lambda u \nabla f$. $\qquad \square$

Proposition 3.11. *Let $\Omega \subset \mathbb{R}^n$ be a convex open set and $f : \Omega \to \mathbb{R}$ be a strictly convex function satisfying the smoothness assumption (SA).*

Then the following statements hold.

(i) If f satisfies Property $\mathcal{F}$ on Ω, then for all non-decreasing, differentiable function $G : \mathbb{R}^+ \to \mathbb{R}^+$, there exists $\varphi \in \mathcal{F}$ such that for $x \in \Omega$ almost everywhere, it holds

$$\nabla \varphi(x) = G(\|\nabla f\|(x)) \frac{\nabla f(x)}{\|\nabla f\|(x)}. \tag{3.12}$$

(ii) Inversely, if there exists a non-decreasing differentiable function $G : \mathbb{R}^+ \mapsto \mathbb{R}^+$ which satisfies $uG'(u) - G(u) \neq 0$ for all $u > 0$, and there exists a convex function $\varphi : \Omega \to \mathbb{R}$ such that equation (3.12) holds almost everywhere, then f satisfies Property $\mathcal{F}$ on Ω.

Proof. The proof of (i) is divided into three steps: we will prove at first the existence of a such function φ satisfying (3.12), secondly we will show that φ is convex, at the end we will prove that $\varphi \in \mathcal{F}$.

Step 1. *Existence of function φ such that (3.12) holds.*

By strict convexity of f, there exists at most one x^* such that $\nabla f(x^*) = 0$.

Denote $\mathbf{F}$ the vector field $G(\|\nabla f\|)\frac{\nabla f}{\|\nabla f\|}$ defined on the simply connected open set $\mathcal{H}_f{}^{\mathrm{o}} \setminus x^*$, then the ith component of $\mathbf{F}$ is :

$$\mathbf{F}_i = G(\|\nabla f\|)\frac{\partial_i f}{\|\nabla f\|}. \tag{3.13}$$

We begin with proving the existence of function φ defined on $\mathcal{H}_f{}^{\mathrm{o}} \setminus \{x^*\}$, according to a generalized version of Poincaré's Lemma [15] and the smoothness assumption (SA), we only need to show that on $\mathcal{H}_f{}^{\mathrm{o}}$, it holds:

$$\mathrm{rot}\ \mathbf{F} = 0,$$

which is equivalent to

$$\partial_j \mathbf{F}_i = \partial_i \mathbf{F}_j \tag{3.14}$$

for all $1 \leqslant i, j \leqslant n$.

Now denote $u = \|\nabla f\|_2$, it holds

$$\partial_j \mathbf{F}_i = \partial_j \left(\frac{G(u)}{u} \right) \partial_i f + \frac{G(u)}{u} \partial_{ij} f$$

$$= \partial_j u \left(\frac{G'(u)u - G(u)}{u^2} \right) \partial_i f + \frac{G(u)}{u} \partial_{ij} f$$

The same argument leads to

$$\partial_i \mathbf{F}_j = \partial_i u \left(\frac{G'(u)u - G(u)}{u^2} \right) \partial_j f + \frac{G(u)}{u} \partial_{ij} f. \tag{3.15}$$

Since $f \in \mathcal{F}$, applying lemma 3.7, we deduce that

$$\partial_j u\ \partial_i f = \partial_i u\ \partial_j f. \tag{3.16}$$

Thus equation (3.14) holds and we have the existence of φ on $\mathcal{H}_f{}^{\mathrm{o}} \setminus \{x^*\}$.

According to the assumption (SA), $\Omega \setminus (\mathcal{H}_f{}^{\mathrm{o}} \setminus \{x^*\})$ is a finite set, together with the fact that Ω is an open set, one can extend the continuous function φ on Ω.

Step 2. φ *is convex.*

We only need to check Hess φ is positive almost everywhere on Ω.

Fix $x \in \mathcal{H}_f{}^{\mathrm{o}} \setminus \{x^*\}$, assume that Hess $f \nabla f = \lambda \nabla f$. It is enough to prove that Hess $\varphi = (\partial_{ij}\varphi)_{ij}$ is positive. According to (3.15) and (3.9), we have

$$\partial_{ij}\varphi = \partial_i \mathbf{F}_j$$

$$= \partial_i u \left(\frac{G'(u)u - G(u)}{u^2} \right) \partial_j f + \frac{G(u)}{u} \partial_{ij} f$$

$$= \frac{G(u)}{u} \left(\partial_{ij} f - \frac{\partial_i u \partial_j f}{u} \right) + \frac{G'(u)}{u^2} \partial_i u\ \partial_j f$$

$$= \frac{G(u)}{u} \left(\partial_{ij} f - \lambda \frac{\partial_i f \partial_j f}{u^2} \right) + \lambda \frac{G'(u)}{u^3} \partial_i f\ \partial_j f.$$

17

Observe that G is positive, non-decreasing and $u > 0$, it follows that $\frac{G(u)}{u} > 0$ and $\frac{G'(u)}{u^3} \geqslant 0$. Moreover, the convexity of f implies that $\lambda \geqslant 0$. Thus, it is enough to prove that the matrix $M_0 = \left(\partial_{ij} f - \lambda \frac{\partial_i f \partial_j f}{u^2} \right)_{ij}$ and $M_1 = (\partial_i f \, \partial_j f)_{ij}$ are positive.

We begin with the positivity of M_1. For any vector $w \in \mathbb{R}^n$,

$$^t w M_1 w = \sum_{i,j} w_i \partial_i f \partial_j f w_j = \left(\sum_i w_i \partial_i f \right)^2 = \langle w, \nabla f \rangle^2 \geqslant 0. \qquad (3.17)$$

Now we turn to prove the positivity of M_0. For any $w \in \mathbb{R}^n$, write $w = y + a\nabla f$ with $a \in \mathbb{R}$ and y perpendicular to ∇f. Noticing that $M_0 = \mathrm{Hess} f - \frac{\lambda}{u^2} M_1$, then

$$^t w M_0 w = {}^t (y + a\nabla f) \left(\mathrm{Hess} f - \frac{\lambda}{u^2} M_1 \right) (y + a\nabla f).$$

Using $\mathrm{Hess} f \, \nabla f = \lambda \nabla f$ and $u^2 = \|\nabla f\|^2$, it holds

$$\begin{aligned}
{}^t (y + a\nabla f) &\mathrm{Hess} f \; (y + a\nabla f) \\
&= \langle y, \mathrm{Hess} f \; y \rangle + a^2 \langle \nabla f, \mathrm{Hess} f \nabla f \rangle + 2a \langle y, \mathrm{Hess} f \nabla f \rangle \\
&= \langle y, \mathrm{Hess} f \; y \rangle + a^2 \langle \nabla f, \lambda \nabla f \rangle + 2a \langle y, \lambda \nabla f \rangle \\
&= \langle y, \mathrm{Hess} f \; y \rangle + a^2 \lambda u^2.
\end{aligned}$$

On the other hand, according to (3.17),

$$^t (y + a\nabla f) \frac{\lambda}{u^2} M_1 (y + a\nabla f) = \frac{\lambda}{u^2} \langle y + a\nabla f, \nabla f \rangle^2 = \frac{\lambda a^2}{u^2} \langle \nabla f, \nabla f \rangle^2 = \lambda a^2 u^2.$$

Thus, together with the convexity of f, we deduce that

$$^t w M_0 w = {}^t (y + a\nabla f) \left(\mathrm{Hess} f - \frac{\lambda}{u^2} M_1 \right) (y + a\nabla f) = \langle y, \mathrm{Hess} f \; y \rangle \geqslant 0.$$

Hence, M_0, M_1 are positive matrices, and $\mathrm{Hess} \, \varphi = \frac{G(u)}{u} M_0 + \lambda \frac{G'(u)}{u^3} M_1$ is positive.

Step 3. $\varphi \in \mathcal{F}$.

The convexity of φ is proved in step 2. It is enough to show that for all $x \in \Omega$ almost everywhere, there exists $\lambda \in \mathbb{R}$ such that $\mathrm{Hess} \varphi(x) \nabla \varphi(x) = \lambda \nabla \varphi(x)$. Adapting the notations in step 1, applying lemma 3.7, it is enough to show that on $\mathcal{H}_f{}^{\mathrm{o}} \setminus \{x^*\}$, it holds

$$\partial_i \|F\| F_j = \partial_j \|F\| F_i. \qquad (3.18)$$

We develop $\partial_i \|F\| F_j$ by (3.13) and (3.15):

$$\begin{aligned}
F_j \partial_i \|F\| &= \frac{G(u)}{u} \partial_j f \frac{1}{\|F\|} \sum_{k=1}^n \frac{G(u)}{u} \partial_k f \left[\partial_i u \left(\frac{G'(u) - G(u)}{u^2} \right) \partial_k f + \frac{G(u)}{u} \partial_i f \right] \\
&= A_{ij} + B_{ij},
\end{aligned}$$

18

where

$$A_{ij} := \frac{G(u)}{u}\partial_j f \frac{1}{\|F\|}\sum_{k=1}^{n}\frac{G(u)}{u}\partial_k f\left[\partial_i u\left(\frac{G'(u)-G(u)}{u^2}\right)\partial_k f\right]$$

and

$$B_{ij} := \frac{G(u)}{u}\partial_j f \frac{1}{\|F\|}\sum_{k=1}^{n}\left(\frac{G(u)}{u}\right)^2\partial_k f\partial_i f.$$

It is easy to see that $B_{ij} = B_{ji}$. Using (3.16), we deduce that $A_{ij} = A_{ji}$. Therefore, equation (3.18) holds and it follows that $\varphi \in \mathcal{F}$. The proof of item (i) is completed.

Now we turn to prove item (ii). Adapting the notations in the step 1 of the proof of item (i), the existence of φ guarantees that

$$\text{rot }\mathbf{F} = 0,$$

Developing the latter equation (see (3.15)), it holds

$$\partial_i u\left(\frac{G'(u)u - G(u)}{u^2}\right)\partial_j f + \frac{G(u)}{u}\partial_{ij}f = \partial_j u\left(\frac{G'(u)u - G(u)}{u^2}\right)\partial_i f + \frac{G(u)}{u}\partial_{ij}f.$$

By assumption of G, we deduce that for almost all $x \in \Omega$,

$$\partial_i u\partial_j f = \partial_j u\partial_i f.$$

The conclusion follows by applying lemma 3.7. $\qquad\square$

Now we present extensions of proposition 2.10 and theorem 2.11 in the case of dimension n.

Proposition 3.19. *Let α, β, θ be differentiable strictly convex cost functions such that $\alpha + \beta = \theta$. We assume that θ is strictly convex. Then if f is differentiable and satisfying Property $\mathcal{F}$ on Ω, there exists $\varphi, \psi \in \mathcal{F}$, such that*

$$f = \varphi + \psi$$

and for all $t > 0$, it holds for $x \in \Omega$ almost surely,

$$Q_t^\theta f(x) = Q_t^\alpha \varphi(x) + Q_t^\beta \psi(x)$$

Proof. Since f satisfying Property $\mathcal{F}$ and (SA) on Ω, according to Proposition 3.11 for $G = \alpha' \circ \theta'^{-1}$, there exists a $\varphi \in \mathcal{F}$ such that

$$\nabla\varphi(x) = \alpha' \circ \theta'^{-1}(\|\nabla f\|(x))\frac{\nabla f(x)}{\|\nabla f\|(x)} \tag{3.20}$$

holds for $x \in \Omega$ almost surely.

Denote $T_t(x)$ the point such that

$$Q_t^\theta f(x) = f(T_t(x)) + t\theta\left(\frac{\|x - T_t(x)\|}{t}\right).$$

19

It follows that $x \in \Omega$ almost surely,

$$\nabla f(T_t(x)) = \theta' \left(\frac{\|x - T_t(x)\|}{t} \right) \frac{x - T_t(x)}{\|x - T_t(x)\|}.$$

Combining equation (3.20), we get

$$\nabla \varphi(T_t(x)) = \alpha' \circ \theta'^{-1}(\|\nabla f\|) \frac{\nabla f}{\|\nabla f\|}(T_t(x))$$

$$= \alpha' \left(\frac{\|x - T_t(x)\|}{t} \right) \frac{x - T_t(x)}{\|x - T_t(x)\|},$$

which implies that for all $x \in \mathbb{R}^n$,

$$Q_t^\alpha \varphi(x) = \varphi(T_t(x)) + t\alpha \left(\frac{\|x - T_t(x)\|}{t} \right). \tag{3.21}$$

Now letting

$$\psi = f - \varphi,$$

it holds

$$\nabla \psi(T_t(x)) = \beta' \left(\frac{\|x - T_t(x)\|}{t} \right) \frac{x - T_t(x)}{\|x - T_t(x)\|}.$$

It follows that

$$Q_t^\alpha \psi(x) = \psi(T_t(x)) + t\beta \left(\frac{\|x - T_t(x)\|}{t} \right). \tag{3.22}$$

Summing (3.21) and (3.22) leads to the conclusion. $\qquad \square$

3.2   Proof of Theorem 3.4

Now we are ready to prove theorem 3.4.

Proof. We first prove (ii): Since g satisfies the smoothness assumption (SA) and

$$x \mapsto g(x) - \frac{1}{2} \langle x, x \rangle$$

is strictly convex, according to Proposition 3.11 by taking $f := x \mapsto g(x) - \frac{1}{2} \langle x, x \rangle$ and $G = \theta'$, there exists a convex function φ such that it holds for $x \in \Omega$ almost everywhere,

$$\nabla \varphi(x) = \theta'(\|x - \nabla g(x)\|) \frac{\nabla g(x) - x}{\|x - \nabla g(x)\|}.$$

Since Ω is the support of μ, it follows that ν almost surely

$$Q^\theta \varphi(\nabla g(x)) = \varphi(x) + \theta(\|x - \nabla g(x)\|).$$

Hence,

$$\mathcal{T}_\theta(\mu, \nu) \leqslant \int \theta(\|x - \nabla g(x)\|)d\nu$$

$$= \int Q^\theta \varphi(\nabla g(x)) - \varphi(x)d\nu$$

$$= \int Q^\theta \varphi d\mu - \int \varphi d\nu \leqslant \overline{\mathcal{T}}_\theta(\nu|\mu).$$

Together with the fact that $\overline{\mathcal{T}}_\theta(\nu|\mu) \leqslant \mathcal{T}_\theta(\mu, \nu)$, the equality holds.

Now we prove that the transfer plan does not depend on the cost function.

Since θ is two times differentiable and convex, applying proposition 3.11 to equation (3.24) for $G = \theta'$, together with the fact that $x \to g(x) - \frac{1}{2}\langle x, x \rangle \in \mathcal{F}$, we deduce that $\varphi \in \mathcal{F}$. Now given a convex cost function α, applying Proposition 3.11 for $G = \alpha' \circ \theta'^{-1}$, we deduce that there exists $\psi \in \mathcal{F}$ such that for ν almost surely,

$$\nabla \psi(x) = \alpha' \circ \theta'^{-1} \frac{\nabla \varphi(x)}{\|\nabla \varphi(x)\|}.$$

It follows that ν almost surely,

$$Q^\alpha \psi(\nabla g(x)) = \psi(x) + \alpha(\|x - \nabla g(x)\|).$$

The conclusion follows by writing the definition of optimal transport and its Kantorovich's duality:

$$\mathcal{T}_\alpha(\mu, \nu) = \inf_\pi \left\{ \int \alpha(\|x - y\|)d\pi \right\} \leqslant \int \alpha(\|x - \nabla g(x)\|)d\nu,$$

and

$$\mathcal{T}_\alpha(\mu, \nu) = \sup_f \left\{ \int Q^\alpha f d\mu - \int f d\nu \right\}$$

$$\geqslant \int Q^\alpha \psi d\mu - \int \psi d\nu$$

$$= \int \alpha(\|x - \nabla g(x)\|)d\nu.$$

By a similar argument with the suprement taken over all f convex, Lipschitz and lower semi-continuous, it holds

$$\overline{\mathcal{T}}_\alpha(\nu|\mu) = \int \alpha(\|x - \nabla g(x)\|)d\nu.$$

We now turn to prove (i): For all convex function φ, it holds for all $x \in \Omega$,

$$Q^\theta \varphi(\nabla g(x)) \leqslant \varphi(x) + \theta(\|x - \nabla g(x)\|). \tag{3.23}$$

Now let φ be the convex function such that $\overline{\mathcal{T}}_\theta(\nu|\mu) = \int Q^\theta \varphi d\mu - \int \varphi d\nu$. Together with (3.23), it follows

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \int Q^\theta \varphi d\mu - \int \varphi d\nu$$
$$\leqslant \int \theta(\|x - \nabla g(x)\|)d\nu = \mathcal{T}_\theta(\mu, \nu).$$

The assumption of (i) implies that the inequality in the latter formula is in fact equality. Thus, for $x \in \Omega$, ν almost surely, equation (3.23) holds. We deduce that

$$\nabla \varphi(x) = \theta'(\|x - \nabla g(x)\|) \frac{\nabla g(x) - x}{\|x - \nabla g(x)\|}. \tag{3.24}$$

According to Proposition 3.11, the conclusion follows.

$\square$

4 Applications

4.1 A simple example

In the space $\mathbb{R}^n$, let μ be the uniform probability measure on the unit ball $B(0, 1)$ and ν be the uniform probability measure on the ball $B(a, 1/4)$. Observe that $T := x \mapsto 4x - a$ satisfies $T \# \nu = \mu$ and define $g(x) := 2\|x\|^2 - ax$ for $x \in B(a, 1/4)$. It is easy to check that $\nabla g = T$ and it is the optimal transfer plan $\mathcal{T}(\mu, \nu)$ with quadratic cost. Now observe that $g(x) - \frac{1}{2}\|x\|^2$ satisfies Property $\mathcal{F}$ on $B(a, 1/4)$. According to item (ii) of theorem 3.4, we can deduce that T is an optimal transfer plan for both weak and classical transport problem with convex costs depending on the distance.

4.2 Links with the infimum convolution inequality

The so-called infimum operator inequalities were first introduced by Maurey in [17]. They are closely related to Transport-cost inequalities.

Let us stay in the space $\mathbb{R}^n$ and adapt the settings before. We say that a probability measure μ satisfies the inf-convolution inequality $\mathrm{IC}(\theta)$ with the cost θ if the following holds for all measurable functions bounded from below $f : \mathbb{R}^n \mapsto \mathbb{R}$:

$$\int e^{Q^\theta f} d\mu \int e^{-f} d\mu \leqslant 1.$$

This inequality was proved to be equivalent to the transport cost inequality (see [5]):

$$\mathcal{T}_\theta(\nu, \mu) \leqslant H(\nu|\mu),$$

where $H(\nu|\mu)$ is the related entropy of ν with respect to μ.

Now consider the inf-convolution inequality restricted to the class $\mathcal{F}$ (denoted by $\mathrm{rIC}(\theta)$):

$$\int e^{Q^\theta f} d\mu \int e^{-f} d\mu \leqslant 1 \qquad \forall f \in \mathcal{F}.$$

22

According to proposition 3.19, let α, β be convex costs such that $\alpha + \beta = \theta$, assume that rIC(α) and rIC(β) hold, then rIC($\frac{1}{2}\theta$) holds. The proof is simply apply Cauchy Schwartz inequality and proposition 3.19, details are left to the readers.

We remark that in one dimension, Property $\mathcal{F}$ is convexity. Feldheim and al. in [9] and Gozlan and al. in [12] proved independently that for a quadratic linear cost α, the inequality rIC(α) is equivalent to the convex Poincaré inequality. For general convex cost θ, rIC(θ) is equivalent to the weak transport inequality $\overline{\mathcal{T}}_\theta \leqslant H$. The fact that rIC($\alpha$) and rIC($\beta$) imply rIC($\theta$) is simply $\overline{\mathcal{T}}_\theta = \overline{\mathcal{T}}_\alpha + \overline{\mathcal{T}}_\beta$ in one dimension.

Acknowledgement

References

[1] A. D. Aleksandrov. Almost everywhere existence of the second differential of a convex functions and related properties of convex surfaces. *Uchenye Zapisky Leningrad. Gos. Univ. Math. Ser.*, 37(37):3–35, 1939.

[2] Julio Backhoff-Veraguas, Mathias Beiglböck, and Gudmund Pammer. Existence, Duality, and Cyclical monotonicity for weak transport costs. *ArXiv e-prints*, (arXiv:1809.05893), 2018.

[3] Julio Backhoff-Veraguas, Mathias Beiglböck, and Gudmund Pammer. Weak monotone rearrangement on the line. *ArXiv e-prints*, (arXiv:1902.05763), 2019.

[4] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84:375–393, 2000.

[5] S. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163(1):1–28, 1999.

[6] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4), 1991.

[7] Lawrence C. Evans. *Partial differential equations*. American Mathematical Society, 2010.

[8] M. Fathi and Y. Shu. Curvature and transport inequalities for markov chains in discrete spaces. *Bernoulli*, 24(1):672–698, 02 2018.

[9] N. Feldheim, A. Marsiglietti, P. Nayar, and J. Wang. A note on the convex infimum convolution inequality. *Bernouilli*, (24(1)), 2015.

[10] W. Gangbo and R. J. McCann. Shape recognition via wasserstein distance. *Q. Appl. Math.*, LVIII(4):705–737, 2000.

[11] N. Gozlan and N. Juillet. On A Mixture Of Brenier and Strassen Theorems. *ArXiv e-prints*, (arXiv:1808.02681), 2018.

[12] N. Gozlan, C. Roberto, P.-M. Samson, Y. Shu, and P. Tetali. Characterization of a class of weak transport-entropy inequalities on the line. *Ann. Inst. H. Poincaré Proba. Statist.*, 54(3):1667–1693, 2018.

[13] N. Gozlan, C. Roberto, P-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *Journal of Functional Analysis*, 273(11):3327 – 3405, 2017.

[14] F. Hirsch, C. Profeta, B. Roynette, and M. Yor. *Peacocks and associated martingales, with explicit constructions*, volume 3 of *Bocconi & Springer Series*. Springer, Milan; Bocconi University Press, Milan, 2011.

[15] S. Mardare. On poincaré and de rham's theorems. *Rev. Roumaine Math. Pures Appl.*, 53, 2008.

[16] A.-W. Marshall, I. Olkin, and B.-C. Arnold. *Inequalities: theory of majorization and its applications*. Springer Series in Statistics. Springer, New York, second edition, 2011.

[17] B. Maurey. Some deviation inequalities. *Geom. Funct. Anal.*, 1(2):188–197, 1991.

[18] R.J. McCann. Existence and uniqueness of monotone measure-preserving maps. *Duke Math. J.*, (163):1–28, 1999.

[19] R. Rado. An inequality. *J. London Math. Soc.*, 27:1–6, 1952.

[20] T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[21] L. Rüschendorf and L. Uckelmann. *On Optimal Multivariate Couplings*. Springer Netherlands, 1997.

[22] Paul-Marie Samson. Transport-entropy inequalities on locally acting groups of permutations. *Electron. J. Probab.*, 22, 2017.

[23] Y. Shu. Hamilton-jacobi equations on graphs and applications. *Potential Analysis*, 48:125–157, 2018.

[24] Y. Shu and M. Strzelecki. Characterization of a class of convex modified log-sobolev inequalities on the line. *Ann. Inst. H. Poincaré Probab. Statist.*, 54(4), 2018.

[25] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965.

[26] C. Villani. *Optimal transport, old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009.