WEAK OPTIMAL TRANSPORT WITH UNNORMALIZED KERNELS

PHILIPPE CHONÉ, NATHAEL GOZLAN AND FRANCIS KRAMARZ

Abstract. We introduce a new variant of the weak optimal transport problem where mass is distributed from one space to the other through unnormalized kernels. We give sufficient conditions for primal attainment and prove a dual formula for this transport problem. We also obtain dual attainment conditions for some specific cost functions. As a byproduct we obtain a transport characterization of the stochastic order defined by convex positively 1-homogenous functions, in the spirit of Strassen theorem for convex domination.

Introduction

The aim of this paper is to study the mathematical aspects of a new variant of the optimal transport problem, related to the weak optimal transport problem introduced in [23], that has been recently considered by the first and third authors in [15] in an economic context.

In what follows $\mathcal{X}$ and $\mathcal{Y}$ are compact metrizable spaces, $\mathcal{P}(\mathcal{X})$ (resp. $\mathcal{P}(\mathcal{Y})$) denotes the set of all Borel probability measures on $\mathcal{X}$ (resp. $\mathcal{Y}$) and $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ are fixed probability measures.

In the usual Monge-Kantorovich transport problem, given a cost function $\omega : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, assumed to be measurable and bounded from below, the optimal transport cost between μ and ν is defined as

$$\mathcal{T}_\omega(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \iint \omega(x,y)\,\pi(dxdy), \tag{1}$$

where $\Pi(\mu,\nu)$ denotes the set of all couplings between μ and ν, that is to say the set of all probability measures π on $\mathcal{X} \times \mathcal{Y}$ such that the $\mathcal{X}$-marginal of π is μ and the $\mathcal{Y}$-marginal of π is ν. We refer to the textbooks [34, 35, 18, 31] for a panorama of applications.

To motivate the introduction of weak optimal transport, recall that any coupling $\pi \in \Pi(\mu,\nu)$ can be disintegrated as follows

$$\pi(dxdy) = \mu(dx)p^x(dy),$$

where $p = (p^x)_{x\in\mathcal{X}}$ is a probability kernel from $\mathcal{X}$ to $\mathcal{Y}$ (which is μ almost surely unique). In an informal way, for all $x \in \mathcal{X}$ the probability $p^x \in \mathcal{P}(\mathcal{Y})$ contains all the information about how the mass taken at x is distributed over $\mathcal{Y}$. Using this notation, one sees in particular that

$$\iint \omega(x,y)\,\pi(dxdy) = \int \left(\int \omega(x,y)\,p^x(dy) \right) \mu(dx),$$

which highlights the fact that in the Monge-Kantorovich optimal transport problem, the mass transfers from $\mathcal{X}$ to $\mathcal{Y}$ are penalized only through their mean costs $\int \omega(x,y)\,p^x(dy)$, $x \in \mathcal{X}$. In contrast, the

Date: March 31, 2022.
2020 *Mathematics Subject Classification.* 60A10 ; 49J55 ; 60G48 ; 90C46.
Key words and phrases. Optimal Transport ; Weak Optimal Transport ; Duality ; Convex order ; Strassen's Theorem.
P. Choné gratefully acknowledges support from the French National Research Agency (ANR), (Labex ECODEC No. ANR-11-LABEX-0047) and F. Kramarz from the European Research Council (ERC), advanced grant FIRMNET (project ID 741467). N. Gozlan is supported by a grant of the Simone and Cino Del Duca Foundation.

Weak Optimal Transport (WOT) framework allows to consider more general penalizations on the probability kernel p. Given a cost function $c : \mathcal{X} \times \mathcal{P}(\mathcal{Y}) \to \mathbb{R}$ assumed to be measurable and bounded from below, the weak optimal transport cost between μ and ν is defined as

$$\mathcal{T}_c(\mu,\nu) = \inf_{p \in \mathcal{P}(\mu,\nu)} \int c(x,p^x)\,\mu(dx), \tag{2}$$

where $\mathcal{P}(\mu,\nu)$ denotes the set of all probability kernels $p = (p^x)_{x\in\mathcal{X}}$ transporting μ onto ν in the sense that

$$\mu p(dy) := \int p^x(dy)\,\mu(dx) = \nu(dy).$$

This definition, which finds its origin in the works by Marton [29, 30] on transport-entropy inequalities and their relations to the concentration of measure phenomenon [21, 26], stricly extends the setting of the Monge-Kantorovich transport problem (which corresponds to $c(x,p) = \int \omega(x,y)\,p(dy)$, $x \in \mathcal{X}$, $p \in \mathcal{P}(\mathcal{Y})$). It turns out that the WOT setting includes several interesting variants of the optimal transport problem such as the Schrödinger / entropic regularized transport problem [27, 16, 10] or the martingale transport problem [8, 19, 9]. General tools such as a Kantorovich type duality formula [23, 2, 4] and a cyclical monotonicity criterium [20, 4] have been developed in the framework of WOT. We refer to the nice survey paper [6] for a general panorama of recent results and applications of WOT. Among the new WOT problems recently considered, the class of barycentric transport problems attracted a particular attention. These barycentric transport problems correspond to cost functions of the form

$$c(x,p) = \theta\left(x - \int y\,p(dy)\right), \qquad x \in \mathcal{X}, p \in \mathcal{P}(\mathcal{Y}),$$

with $\mathcal{X},\mathcal{Y} \subset \mathbb{R}^d$ and $\theta : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ a convex function bounded from below. The introduction of these barycentric optimal transport costs was first motivated by their applications in concentration of measure [23, 22]. In dimension 1 and for a general convex function θ, the structure of optimal plans has been settled in [22, 1, 5]. For the quadratic cost function $\theta = \|\cdot\|_2^2$ on the Euclidean space $(\mathbb{R}^d, \|\cdot\|_2)$ with $d \geqslant 1$, the structure of optimal plans has been described in [20, 4], and yields a new characterization of the couples (μ,ν) for which the Brenier transport map [11, 12] is a contraction and to a new proof [17] of the Caffarelli contraction theorem [13]. These barycentric cost functions also recently found applications in machine learning [14].

The new variant of the WOT problem studied in the present paper consists in relaxing the assumption that $p = (p^x)_{x\in\mathcal{X}}$ appearing in (2) is a *probability* kernel. To state a formal definition, we need to introduce additional notions and notations. We will denote by $\mathcal{M}(\mathcal{Y})$ the set of all finite nonnegative measures on $\mathcal{Y}$. This set will always be equipped with the usual weak topology, and with the cylindric σ-field. A *nonnegative kernel from $\mathcal{X}$ to $\mathcal{Y}$* is a collection $q = (q^x)_{x\in\mathcal{X}}$ of elements of $\mathcal{M}(\mathcal{Y})$ such that the map $\mathcal{X} \to \mathbb{R}_+ : x \mapsto q^x(A)$ is measurable for all Borel set $A \subset \mathcal{Y}$. Given $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ and a measurable cost function $c : \mathcal{X} \times \mathcal{M}(\mathcal{Y}) \to \mathbb{R}$ such that there exist $r_0, r_1 \in \mathbb{R}$ such that

$$c(x,m) \geqslant r_0 + r_1 m(\mathcal{Y}), \qquad \forall x \in \mathcal{X}, \forall m \in \mathcal{M}(\mathcal{Y}), \tag{LB}$$

we consider

$$\mathcal{I}_c(\mu,\nu) = \inf_{q \in \mathcal{Q}(\mu,\nu)} \int c(x,q^x)\,\mu(dx), \tag{3}$$

where $\mathcal{Q}(\mu,\nu)$ denotes the set of all nonnegative kernels from $\mathcal{X}$ to $\mathcal{Y}$ such that $\mu q = \nu$, where as above $\mu q(dy) = \int q^x(dy)\,\mu(dx)$.

Remark 0.1. *The same transport problem can be stated for measures μ, ν with different masses. It is not difficult to see that this unbalanced problem can be reduced to the one above simply by redefining the cost function. So in all the paper, we will stick to the balanced case.*

The transport problem (3) has been introduced in [15] with the following economic motivation. The space $\mathcal{X}$ represents firms' technologies in a given industry and the space $\mathcal{Y}$ represents workers' skills. The probability measures μ and ν represent the distributions of firms and of workers respectively. A firm of type $x \in \mathcal{X}$ that recruits a distribution of workers $m \in \mathcal{M}(\mathcal{Y})$ produces output given by $-c(x, m)$. The problem (3) consists in maximizing total output in the industry over all possible assignments of workers to firms. More precisely, a nonnegative kernel $q \in \mathcal{Q}(\mu, \nu)$ represents a particular hiring policy, with $q^x(dy)$ giving the distribution of workers hired by firms of type $x \in \mathcal{X}$. The condition $\mu q = \nu$ expresses that all workers are employed. The mass $N(x) := q^x(\mathcal{Y})$ represents the total number of workers hired by a firm $x \in \mathcal{X}$, i.e., the size of firms with technology x. Importantly, these firms' sizes are an outcome of the optimization process, whereas OT models restrict to *probability* kernels and hence cannot accommodate this issue.

The main difficulty in dealing with the transport problem (3) is that, unlike Problem (2), assuming that the cost function is jointly lower semicontinuous and convex in its second variable is not enough to ensure existence of a minimizer. To obtain existence of a minimizer, one needs to introduce additional conditions :

- We will say that the cost function $c : \mathcal{X} \times \mathcal{M}(\mathcal{Y}) \to \mathbb{R}$ satisfies Assumption (A) if there exists a family of continuous functions $(a_k)_{k\geqslant 0}$ on $\mathcal{X}$ and a family of continuous functions $(b_k)_{k\geqslant 0}$ on $\mathcal{X} \times \mathcal{Y}$ such that

$$c(x, m) = \sup_{k\geqslant 0} \left\{ \int b_k(x, y)\, m(dy) + a_k(x) \right\}, \qquad x \in \mathcal{X}, m \in \mathcal{M}(\mathcal{Y}). \tag{A}$$

  Note that this condition implies in particular that c is jointly lower semicontinuous, convex with respect its second variable and satisfies (LB).
- We will say that c satisfies Assumption (B) if

$$\frac{c(x, \lambda m)}{\lambda} \underset{\lambda\to\infty}{\longrightarrow} +\infty, \qquad \forall x \in \mathcal{X},\ \forall m \in \mathcal{M}(\mathcal{Y})\backslash\{0\}. \tag{B}$$

Let us now present the main results of this paper.

Our first main contribution, is a primal attainment result for the transport problem (3). Under Assumptions (A) and (B), we show that for all probability measures $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$, there exists a nonnegative kernel $q \in \mathcal{Q}(\mu, \nu)$ such that $\mathcal{I}_c(\mu, \nu) = \int c(x, q^x)\, \mu(dx)$ (see Theorem 2.2). In a nutshell, the role of Assumption (B) is to avoid mass accumulation on sets of μ measure 0. Note that existence of solutions can also hold under other types of conditions on c (see in particular Theorem 5.4 dealing with nonpositive cost functions c having a moderate growth).

Our second main result is a Kantorovich type duality formula for the transport problem (3). Under assumptions (A) and (B), for all probability measures $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$, it holds

$$\mathcal{I}_c(\mu, \nu) = \sup_{f\in\mathcal{C}_b(\mathcal{Y})} \left\{ \int K_c f(x)\, \mu(dx) - \int f(y)\, \nu(dy) \right\},$$

where $\mathcal{C}_b(\mathcal{Y})$ denotes the set of (bounded) continuous functions on $\mathcal{Y}$ and the operator K_c is defined by

$$K_c f(x) = \inf_{m\in\mathcal{M}(\mathcal{Y})} \left\{ \int f\, dm + c(x, m) \right\}, \qquad x \in \mathcal{X}.$$

Note that, at least formally, if one allows c to take the value $+\infty$ and $c(\,\cdot\,, m) = +\infty$ when m is not of mass 1, then $\mathcal{I}_c(\mu, \nu) = \mathcal{T}_c(\mu, \nu)$ and one recovers the duality formula for WOT [4]. As we shall see in Theorem 3.2, the duality formula for $\mathcal{I}_c$ actually holds under a more general condition (Approx) which is in particular implied by Assumption (B).

The third main contribution of this paper is a general investigation of transport problems involving cost functions of the following form

$$c(x,m) = F\left(x, \int y\, dm\right), \qquad x \in \mathcal{X}, m \in \mathcal{M}(\mathcal{Y}) \tag{4}$$

where $\mathcal{Y}$ is a compact subset of $\mathbb{R}^d$ whose conical hull is denoted by $\mathcal{Z}$ and $F : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$ is convex with respect to its second variable. Such cost functions will be called *conical* in all the paper. This name comes from the fact that cost functions of this type are naturally related to *positively* 1-*homogenous convex* functions (whose epigraphs are cones). Recall that a function $\varphi : \mathbb{R}^d \to \mathbb{R}$ is said positively 1-homogenous (or positively homogenous of degree 1) if $\varphi(tx) = t\varphi(x)$ for all $t \geqslant 0$ and $x \in \mathbb{R}^d$. This link between conical cost functions and positively 1-homogenous convex appears in the duality formula for the transport problem (3). More precisely, we will prove that if c is a conical cost function satisfying (A) and such that the convex hull of the support of ν does not contain 0, then, under some mild integrability condition on F, the following reduced duality formula holds

$$\mathcal{I}_c(\mu,\nu) = \sup_{\varphi} \left\{ \int Q_F\varphi(x)\, \mu(dx) - \int \varphi(y)\, \nu(dy) \right\}, \tag{5}$$

where φ runs over the set of positively 1-homogenous convex functions and the operator Q_F is defined by

$$Q_F\varphi(x) = \inf_{z\in\mathcal{Z}} \{\varphi(z) + F(x,z)\}, \qquad x \in \mathcal{X}.$$

Moreover, under the same assumptions, we will show the existence of *dual optimizers* ; see Theorem 5.1 for a precise statement. These conical cost functions are precisely those that were considered in [15] and motivated the present paper. In the economic model of [15], a dual optimizer φ represents a "wage schedule" ($\varphi(y)$ is the wage paid to workers with skills y), while $Q_F\varphi(x)$ represents the opposite of the profit earned by firms with technology x (the profit is the produced output $-F(x,z)$ minus the firm's wage bill $\varphi(z)$, with z being the sum of the skills of firm x' employees).

A byproduct of our primal attainment and duality results for conical cost functions is a seemingly new variant of the Strassen's theorem [33], that we shall now present. Recall that if μ,ν are probability measures on $\mathbb{R}^d$ having finite first moments, one says that μ is dominated by ν in the convex order, which is denoted by $\mu \leqslant_c \nu$, if

$$\int f\, d\mu \leqslant \int f\, d\nu \tag{6}$$

for all convex function $f : \mathbb{R}^d \to \mathbb{R}$. Strassen's theorem provides the following useful probabilistic characterization of convex order : $\mu \leqslant_c \nu$ if and only if there exists a couple of random variables (X_0, X_1) such that $X_0 \sim \mu$, $X_1 \sim \nu$ and (X_0, X_1) is a martingale :

$$\mathbb{E}[X_1 \mid X_0] = X_0 \qquad \text{a.s.}$$

Note that if $\pi(dxdy) = \mu(dx)p^x(dy)$ denotes the law of (X_0, X_1), the martingale condition is equivalent to the following centering condition on the probability kernel p : for μ almost all x,

$$\int y\, p^x(dy) = x.$$

Our generalization of Strassen's theorem deals with a weaker variant of the convex order defined as follows : if (6) holds for all positively 1-homogenous convex functions f, we will say that μ is dominated by ν in the *positively* 1-*homogenous convex order* and write $\mu \leqslant_{phc} \nu$. As we will see in Theorem 5.2, if ν is a compactly supported probability measure on $\mathbb{R}^d$ such that the convex hull of the support of ν does not contain 0, then $\mu \leqslant_{phc} \nu$ if and only if there exists a nonnegative kernel

$q \in \mathcal{Q}(\mu,\nu)$ such that, for μ almost all x,

$$\int y\, q^x(dy) = x. \tag{7}$$

See Theorem 5.2 for the case where 0 belongs to the convex hull of the support of μ. Let us briefly explain how this Strassen type result is connected to conical costs. Consider μ, ν two compactly supported probability measures on $\mathbb{R}^d$ and denote by $\mathcal{Y}$ the support of ν. For $p > 1$, the conical cost function

$$c(x,m) = \left\| x - \int y\, dm \right\|^p, \qquad x \in \mathbb{R}^d, m \in \mathcal{M}(\mathcal{Y}),$$

where $\|\cdot\|$ is an arbitrary norm on $\mathbb{R}^d$ satisfies Assumption (B) if and only if 0 does not belong to the convex hull of $\mathcal{Y}$. According to our primal existence result, we thus have $\mathcal{I}_c(\mu,\nu) = 0$ if and only if there is some $q \in \mathcal{Q}(\mu,\nu)$ satisfying (7). Using the dual formulation (5), one can then show with some extra work, that $\mu \leqslant_{phc} \nu$ implies that $\mathcal{I}_c(\mu,\nu) = 0$, thus completing the proof. In our proof, we actually follow a slightly different route, since we use the cost c above with $p = 1$ which will allow us to relax the assumption on the support of ν.

The new version of Strassen's theorem will also enable us to describe optimal transport plans for conical transport costs, in the spirit of [20]. As we will see in Theorem 5.6, as soon a conical cost function c of the form (4) satisfies Assumption (A) and the convex hull of the support of ν does not contain 0, it holds

$$\mathcal{I}_c(\mu,\nu) = \inf_{\gamma \leqslant_{phc} \nu} \mathcal{T}_F(\mu,\gamma), \tag{8}$$

where $\mathcal{T}_F$ denotes the Monge-Kantorovich optimal transport cost associated to the cost function F:

$$\mathcal{T}_F(\mu,\gamma) = \inf_{\pi \in \Pi(\mu,\gamma)} \iint F(x,z)\, \pi(dxdz), \qquad \forall \mu \in \mathcal{P}(\mathcal{X}), \forall \gamma \in \mathcal{P}(\mathcal{Z}).$$

Moreover, if q is a kernel minimizer for $\mathcal{I}_c(\mu,\nu)$, then the map $S(x) = \int y\, q^x(dy)$, $x \in \mathcal{X}$, does not depend on the particular choice of the optimizer $\bar{q}$ and provides an optimal transport for the cost $\mathcal{T}_F$ between μ and a probability measure $\bar{\nu} \leqslant_{phc} \nu$ that achieves the infimum in (8). The map S can also be related to dual optimizers (see Theorem 5.7 and Corollary 5.1). In the particular case where $\mathcal{X}$ is a compact subset of $\mathbb{R}^d$ and $F(x,z) = \frac{1}{2}\|x-z\|_2^2$, $x, z \in \mathbb{R}^d$, with $\|\cdot\|_2$ the standard Euclidean norm, more can be said about the form of the transport map S. Namely, we show in Theorem 5.8 that there exists some closed convex set C such that for μ almost every x, it holds $S(x) = x - p_C(x)$ where p_C is the orthogonal projection onto the set C.

Let us point out that during the preparation of this work, we learned about the recent paper [25], devoted to the study of

$$\inf_{\gamma \leqslant_{\mathcal{A}} \nu} \mathcal{T}_F(\mu,\gamma),$$

where $F : \mathcal{X} \times \mathbb{R}^d \to \mathbb{R}_+$ is some lower semicontinuous function, $\mathcal{A}$ is some cone of continuous functions on $\mathbb{R}^d$ and $\gamma \leqslant_{\mathcal{A}} \nu$ means that $\int f\, d\gamma \leqslant \int f\, d\nu$ for all $f \in \mathcal{A}$. A general duality formula has been obtained in [25] for these distance functionals (called backward projection there) : under good assumptions, it holds

$$\inf_{\gamma \leqslant_{\mathcal{A}} \nu} \mathcal{T}_F(\mu,\gamma) = \sup_{\varphi \in \mathcal{A}} \int Q_F\varphi\, d\mu - \int \varphi\, d\nu,$$

with Q_F defined as above (with $\mathcal{Z} = \mathbb{R}^d$). We refer to [25, Theorem 4.3] for a precise statement. Applying this result to the class $\mathcal{A}$ of all convex positively 1-homogenous functions together with (8), gives back the duality formula (5). Note that the two papers complement each other, since the identity (8) crucially requires the variant of Strassen theorem for the convex positively 1-homogenous order proved here. It would be very interesting to see if other forward projections admit representations in terms of weak transport costs $\mathcal{T}_c$ or $\mathcal{I}_c$ for some special classes of cost functions c, but this question will not be considered here.

In the conical case described above, a basic feature of the corresponding transport problem is that it admits in general more than one solution. This non-uniqueness of solutions is no longer true for other class of cost functions, also considered in [15], that we shall now describe. Suppose that $c : \mathcal{X} \times \mathcal{M}(\mathcal{Y}) \to \mathbb{R}$ is given by

$$c(x, m) = G\left(\int F(x, y)\, m(dy)\right), \qquad x \in \mathcal{X}, m \in \mathcal{M}(\mathcal{Y}), \tag{9}$$

where $F : \mathcal{X} \times \mathcal{Y} \to (0, \infty)$ is some continuous function and $G : [0, \infty) \to \mathbb{R}$ is a convex differentiable function. We establish in Theorem 4.2, that when $\mathcal{X}, \mathcal{Y}$ are compact subsets of $\mathbb{R}$, μ has no atoms, G is monotonic and $F : \mathbb{R}^2 \to \mathbb{R}$ is twice continuously differentiable and satisfies the following condition

$$\frac{\partial \ln F}{\partial x \partial y} \neq 0$$

then the transport problem (3) associated to a cost function of the form (9) admits at most one kernel solution (and exactly one whenever $G'(x) \to +\infty$ as $x \to \infty$ for instance). Moreover, this kernel solution is of the following form

$$\bar{q}^x(dy) = \bar{N}(x)\delta_{\bar{T}(x)}(dy),$$

for μ almost every $x \in \mathcal{X}$, where $\bar{N}$ is a density with respect to μ and $\bar{T}$ is a monotonic function. This uniqueness result is obtained as a consequence of a general result of independent interest establishing a relation between the support of primal solutions and dual optimizers (see Proposition 4.1 for details).

The paper is organized as follows. In Section 1, we introduce another formulation of the transport problem (3) involving couplings π with a first marginal absolutely continuous with respect to μ and second marginal equal to ν. This class of couplings being not closed in general, primal attainment is not always true (when it holds we call such coupling a strong solution). To compensate this non-attainment issue, we introduce the notion of *weak solution*. These weak solutions are defined as limit points of minimizing sequences, and as such always exist. We conclude Section 1 by giving several explicit examples admitting only strong solutions or only weak (but not strong) solutions or solutions of both types. In Section 2, we show that under good assumptions, weak solutions can be interpreted as couplings minimizing a certain functional denoted $\bar{I}_c^{\mu}$ which is lower semicontinuous on its domain of definition. One of the main result of this section is Theorem 2.2 which shows that under Assumption (B) all weak solutions are strong. The main result of Section 3, Theorem 3.2, provides the dual formulation of the transport problem already presented above. Section 4 deals with cost functions of the form (9). We prove in Theorem 4.1 that the dual problem admits at least one solution. Then we establish in Proposition 4.1 a general link between supports of primal solutions and this dual optimizer, on which relies the proof of the uniqueness result (Theorem 4.2) presented above. Section 5 is entirely devoted to the study of the transport problem (3) for conical cost functions. We prove in particular in Theorem 5.1 duality and dual attainment under conditions that are weaker than in Theorem 3.2. This section also contains the Strassen type result presented above characterizing the positively 1-homogenous convex order (Theorem 5.2) and, as a corollary, the identity (8). We also prove in Theorem 5.4 a primal attainment result for a special class of nonpositive conical cost functions. Finally, the paper ends with an Appendix containing the proofs of some technical results of Section 2.

Contents

1. A new transport problem

In this section, we first introduce an alternative equivalent formulation of the transport problem (3) which involves couplings and is thus closer to the usual optimal transport framework. Then we introduce the notion of weak solutions which compensate the fact that the transport problem (3) does not always admit minimizers. Finally, we study several explicit examples of transport problem (3) in dimension one.

1.1. Definitions, equivalent formulation and first properties.

First let us introduce some notations. If E is some Polish metric space, we will denote by $\mathcal{P}(E)$ the set of all Borel probability measures on E and by $\mathcal{M}(E)$ (resp. $\mathcal{M}_s(E)$) the set of all nonnegative finite measures (resp. finite signed measures) on E. The space $\mathcal{M}_s(E)$ will be equipped with the topology of weak convergence, that is to say the coarsest topology that makes the maps $\mathcal{M}_s(E) \to \mathbb{R} : m \mapsto \int f\, dm$ continuous for all $f \in \mathcal{C}_b(E)$, the space of all bounded continuous functions on E.

In all what follows, $\mathcal{X}$ and $\mathcal{Y}$ will be two compact metrizable spaces and $c : \mathcal{X} \times \mathcal{M}(\mathcal{Y}) \to \mathbb{R}$ will be a cost function which will always be assumed to be convex with respect to its second variable, jointly lower semicontinuous on $\mathcal{X} \times \mathcal{M}(\mathcal{Y})$ and to satisfy the lower bound (LB). Given $\mu \in \mathcal{P}(\mathcal{X})$, for any nonnegative kernel q from $\mathcal{X}$ to $\mathcal{Y}$ such that $\mu q(\mathcal{Y}) = 1$ we will set

$$I_c^\mu[q] = \int c(x, q^x)\, \mu(dx).$$

With this notation, the transport problem (3) can be restated as

$$\mathcal{I}_c(\mu, \nu) = \inf_{q \in \mathcal{Q}(\mu,\nu)} I_c^\mu[q], \tag{10}$$

where we recall that $\mathcal{Q}(\mu, \nu)$ denotes the set of all nonnegative kernels q from $\mathcal{X}$ to $\mathcal{Y}$ such that $\mu q = \nu$. We will say that $\bar{q} \in \mathcal{Q}(\mu, \nu)$ is a *kernel solution* to the transport problem (10) if

$$\mathcal{I}_c(\mu, \nu) = \int c(x, \bar{q}^x)\, \mu(dx).$$

A first basic observation is that $\mathcal{I}_c$ is jointly convex.

Proposition 1.1. *The functional* $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y}) \to \mathbb{R} \cup \{+\infty\} : (\mu, \nu) \mapsto \mathcal{I}_c(\mu, \nu)$ *is convex.*

Proof. Take $\mu_0, \mu_1 \in \mathcal{P}(\mathcal{X})$, $\nu_0, \nu_1 \in \mathcal{P}(\mathcal{Y})$ and for $t \in]0, 1[$ let $\mu_t = (1-t)\mu_0 + t\mu_1$ and $\nu_t = (1-t)\nu_0 + t\nu_1$. It will be convenient to work with a reference probability measure m such that $\mu_t \ll m$ for all $t \in [0, 1]$. One can take for instance $m = \mu_{1/2}$, since $\mu_{1/2}(A) = 0$ implies that $\mu_0(A) = \mu_1(A) = 0$ which implies that $\mu_t(A) = 0$. Let q_0 and q_1 be nonnegative kernels such that $\mu_0 q_0 = \eta_0$ and $\mu_1 q_1 = \eta_1$. Let q_t the nonnegative kernel defined by

$$q_t^x(dy) = \frac{(1-t)h_0(x)}{(1-t)h_0(x) + th_1(x)} q_0^x(dy) + \frac{th_1(x)}{(1-t)h_0(x) + th_1(x)} q_1^x(dy),$$

where h_0 and h_1 are the densities of μ_0, μ_1 with respect to m. We have

$$\begin{aligned}\mu_t q_t(dy) &= \int q_t^x(dy)\mu_t(dx) = \int [(1-t)h_0(x)q_0^x(dy) + th_1(x)q_1^x(y)]\, m(dx)\\ &= (1-t)\nu_0(dy) + t\nu_1(dy) = \nu_t(dy),\end{aligned}$$

hence $\mu_t q_t = \nu_t$. By convexity of $c(x, \cdot)$, it holds

$$\begin{aligned}\int c(x, q_t^x)\, \mu_t(dx) &\leqslant \int (1-t)h_0(x)c(x, q_0^x) + th_1(x)c(x, q_1^x)\, m(dx)\\ &= (1-t)\int c(x, q_0^x)\, \mu_0(dx) + t\int c(x, q_1^x)\, \mu_1(dx).\end{aligned}$$

The result then follows by minimizing over q_0 and q_1. □

Let us now derive an alternative equivalent formulation of the transport problem (10) that is closer to the classical viewpoint in optimal transport. We will denote by $\Pi(\eta, \nu)$ the set of all transport plans between two probability measures $\eta \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$, that is to say the set of probability measures on $\mathcal{X} \times \mathcal{Y}$ having η and ν as marginals. For any $\mu \in \mathcal{P}(\mathcal{X})$, we will consider

$$\Pi(\ll \mu, \nu) = \bigcup_{\eta \ll \mu} \Pi(\eta, \nu) \qquad \text{and} \qquad \Pi(\ll \mu, \cdot\,) = \bigcup_{\nu \in \mathcal{P}(\mathcal{Y})} \Pi(\ll \mu, \nu)$$

where $\eta \ll \mu$ means that η is absolutely continuous with respect to μ. In other words, $\Pi(\ll \mu, \cdot\,)$ is the set of all probability measures on $\mathcal{X} \times \mathcal{Y}$ whose first marginal is absolutely continuous with respect to μ. Observe that if $q \in \mathcal{Q}(\mu, \nu)$ then the function N defined by $N(x) = q^x(\mathcal{Y})$ is such that

$$\int N(x)\, \mu(dx) = 1.$$

Therefore, N is a probability density with respect to μ. Moreover, $\pi(dxdy) = \mu(dx)q^x(dy)$ is a transport plan between $\eta(dx) := N(x)\,\mu(dx)$ and $\nu(dy)$. Conversely, if $\eta \in \mathcal{P}(\mathcal{X})$ is absolutely continuous with respect to μ and $\pi \in \Pi(\eta, \nu)$ with $\pi(dxdy) = \eta(dx)p^x(dy)$, then the nonnegative kernel q defined by $q^x(dy) = \frac{d\eta}{d\mu}(x)p^x(dy)$, $x \in \mathcal{X}$, belongs to $\mathcal{Q}(\mu, \nu)$. With a slight abuse of notation, let us also denote by I_c^μ the function defined on $\Pi(\ll \mu, \cdot)$ by

$$I_c^\mu[\pi] = \int c\left(x, \frac{d\pi_1}{d\mu}(x)p^x\right)\mu(dx), \qquad \pi \in \Pi(\ll \mu, \cdot),$$

where π_1 is the first marginal of π and p is the probability kernel such that $\pi(dxdy) = \pi_1(dx)p^x(dy)$. With this notation, it thus holds

$$\mathcal{I}_c(\mu, \nu) = \inf_{\pi \in \Pi(\ll \mu, \nu)} I_c^\mu[\pi]. \tag{11}$$

Definition 1.1 (Strong solutions)**.** *Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$; a probability measure $\bar\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ is called a* strong *solution of the transport problem* (10) *if $\bar\pi \in \Pi(\ll \mu, \nu)$ and $\mathcal{I}_c(\mu, \nu) = I_c^\mu[\bar\pi]$.*

Note that if $\bar q$ is a kernel solution to the transport problem (10) then the transport plan $\bar\pi(dxdy) = \mu(dx)\bar q^x(dy) \in \Pi(\ll \mu, \nu)$ is a strong solution of the transport problem (10) and, conversely, any strong solution defines a kernel solution.

Since the set $\Pi(\ll \mu, \nu)$ is not closed in general, the infimum in the transport problem (10) is not always attained and strong solutions may not always exist. This technical issue motivates the introduction of weak solutions.

Definition 1.2 (Weak solutions)**.** *Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$; a probability measure $\bar\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ is called a* weak *solution of the transport problem* (10) *if there exists a sequence of transport plans $\pi_n \in \Pi(\ll \mu, \nu)$ such that $\pi_n \to \bar\pi$ for the weak topology and $I_c^\mu[\pi_n] \to \mathcal{I}_c(\mu, \nu)$.*

Of course, a strong solution is also a weak solution. Under good conditions on the cost function c, weak solutions will be interpreted in Section 2.2 as solutions of a related minimization problem.

Weak solutions always exist as shows the following elementary result.

Proposition 1.2. *For any $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$, the transport problem* (10) *admits at least one weak solution.*

Proof. Let $(\pi_n)_{n\in\mathbb{N}}$ be a minimizing sequence in $\Pi(\ll \mu, \nu)$, that is to say that $\lim_{n\to\infty} I_c^\mu[\pi_n] = \mathcal{I}_c(\mu, \nu)$. Since $\mathcal{X} \times \mathcal{Y}$ is compact, the space $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ is also compact. Therefore, the sequence $(\pi_n)_{n\in\mathbb{N}}$ admits at least one converging subsequence, and any limit point $\bar\pi$ is a weak solution of the transport problem (10). □

At least in the simple case when $\mathcal{X}$ is a finite set, strong solutions always exist.

Theorem 1.1. *Suppose that $\mathcal{X}$ is a finite then, for any $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$, every weak solution of the transport problem* (10) *is a strong solution.*

We will need the following lemma.

Lemma 1.1. *If $\mathcal{X}$ is a finite set, then for any $\mu \in \mathcal{P}(\mathcal{X})$ the set $\Pi(\ll \mu, \cdot)$ is a closed subset of $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ and the functional $\Pi(\ll \mu, \cdot) \to \mathbb{R} : \pi \mapsto I_c^\mu[\pi]$ is lower semicontinuous.*

Proof. Fix $\mu \in \mathcal{P}(\mathcal{X})$. The fact that $\Pi(\ll \mu, \cdot)$ is a closed subset of $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ is easy to see and left to the reader. Let $(\pi_n)_{n\in\mathbb{N}}$ be a sequence of $\Pi(\ll \mu, \cdot)$ converging to some π. Since $\Pi(\ll \mu, \cdot)$ is closed, it follows that π belongs to $\Pi(\ll \mu, \cdot)$. Denote by η_n (resp. η) the first marginal of π_n

(resp. π). For all $x \in \mathcal{X}$ such that $\eta_n(x) > 0$, $\pi_n^x(dy) = \frac{\pi_n(x,dy)}{\eta_n(x)}$. If $\eta_n(x) = 0$, set $\pi_n^x(dy) = \nu(dy)$ (say). Then $\eta_n \to \eta$ and $\frac{d\eta_n}{d\mu}(x) \to \frac{d\eta}{d\mu}(x)$ for all $x \in \mathcal{X}$ such that $\mu(x) > 0$. Also, it is clear that $\pi_n^x(dy) = \frac{\pi_n(x,dy)}{\eta_n(x)} \to \frac{\pi(x,dy)}{\eta(x)} = \pi^x(dy)$ as $n \to \infty$, for all x such that $\eta(x) > 0$. So, using the lower semicontinuity of c, one gets

$$\begin{aligned}\liminf_{n\to\infty} I_c^\mu[\pi_n] = \liminf_{n\to\infty} \sum_{x\in\mathcal{X}} c\left(x, \frac{d\eta_n}{d\mu}(x)\pi_n^x\right)\mu(x) &\geqslant \sum_{x\in\mathcal{X}} \liminf_{n\to\infty} c\left(x, \frac{d\eta_n}{d\mu}(x)\pi_n^x\right)\mu(x)\\ &\geqslant \sum_{x\in\mathcal{X}} c\left(x, \frac{d\eta}{d\mu}(x)\pi^x\right)\mu(x) = I_c^\mu[\pi],\end{aligned}$$

which completes the proof. □

Proof of Theorem 1.1. Let π be some weak solution of the transport problem (10) and $(\pi_n)_{n\in\mathbb{N}}$ be a minimizing sequence converging to π. Since $\mathcal{X}$ is finite, it follows from Lemma 1.1 that $\Pi(\ll\mu,\nu)$ is closed, and so $\pi \in \Pi(\ll\mu,\nu)$. According to Lemma 1.1, it follows that $\mathcal{I}_c(\mu,\nu) = \liminf_{n\to\infty} I_c^\mu[\pi_n] \geqslant I_c^\mu[\pi]$, and so π is a strong solution. □

1.2. **Examples.** We study below particular cases of the transport problem (10) and we describe their set of solutions. These explicit examples show that all the possibility in terms of uniqueness or non-uniqueness or existence of strong solutions can occur.

1.2.1. *An example without strong solution.* Suppose that μ is the uniform measure on $\mathcal{X} = [0,1]$ and ν is an arbitrary probability measure on $\mathcal{Y} = [2,3]$ and define

$$c(x,m) = \int |x-y|^2\, m(dy), \quad x \in [0,1], \qquad m \in \mathcal{M}(\mathcal{Y}).$$

Then,

$$\mathcal{I}_c(\mu,\nu) = \inf_{\mu q=\nu} \iint |y-x|^2 \mu(dx) q^x(dy).$$

Since for all $x \in [0,1]$ and $y \in [2,3]$, $|y-x|^2 \geqslant |y-1|^2$ it holds

$$\mathcal{I}_c(\mu,\nu) \geqslant \int_2^3 |y-1|^2\,\nu(dy).$$

This lower bound is not reached. Indeed, suppose by contradiction that there is some $q \in \mathcal{Q}(\mu,\nu)$ such that $\iint |y-x|^2\mu(dx)q^x(dy) = \int |y-1|^2\,\nu(dy)$. Then, denoting by $\pi(dxdy) = \mu(dx)q^x(dy)$ the associated transport plan, it would hold $\pi(\{1\}\times[2,3]) = 1$ and so $\pi = \delta_1 \otimes \nu$ and $\pi_1 = \delta_1$. Since δ_1 is not absolutely continuous with respect to μ this is not possible. So this problem does not admit strong solutions.

On the other hand, define for all $n \geqslant 2$, $\pi_n(dxdy) = \eta_n(dx)\otimes\nu(dy)$, with η_n the uniform probability measure on $[1-1/n, 1]$. The associated kernel is given by

$$q_n^x(dy) = n\mathbf{1}_{[1-1/n,1]}(x)\nu(dy), \qquad x \in [0,1],$$

and it holds

$$\begin{aligned}\int c(x,q_n^x)\,\mu(dx) &= \int_0^1\int_2^3 |y-x|^2 q_n^x(dy)\mu(dx) = n\int_2^3\int_{1-1/n}^1 |y-x|^2 dx\nu(dy)\\ &\leqslant \int_2^3 |y-(1-1/n)|^2\,\nu(dy) \to \int_2^3 |y-1|^2\,\nu(dy)\end{aligned}$$

as $n\to\infty$. This shows that $\mathcal{I}_c(\mu,\nu) = \int_2^3 |y-1|^2\,\nu(dy)$ and that $\pi = \delta_1\otimes\nu$ is a weak solution of this transport problem.

1.2.2. *An example with a unique strong solution.* In this paragraph, we modify the definition of the cost function of the first example and observe the effect in terms of existence of strong solutions. Let μ be the uniform distribution on $\mathcal{X} = [0,1]$ and $\nu = \delta_2$ on $\mathcal{Y} = [2,3]$ and consider now the cost

$$c(x,m) = \left(\int |y-x|\, dm\right)^2, \quad x \in [0,1], \qquad m \in \mathcal{M}(\mathcal{Y}).$$

If $q \in \mathcal{Q}(\mu,\delta_2)$ then $q^x(\mathbb{R}\backslash\{2\}) = 0$ for almost all x. Therefore, denoting $N(x) = q^x(\{2\})$, it holds

$$I_c^\mu[q] = \int_0^1 (2-x)^2 N^2(x)\, dx.$$

By Cauchy-Schwarz,

$$1 = \int_0^1 N(x)\, dx \leqslant \left(\int_0^1 (2-x)^2 N^2(x)\, dx\right)^{1/2} \left(\int_0^1 \frac{1}{(2-x)^2}\, dx\right)^{1/2}.$$

So, letting $C = \left(\int_0^1 \frac{1}{(2-x)^2}\, dx\right)^{-1}$, we get that

$$I_c^\mu[q] \geqslant C$$

and there is equality if and only if $N(x) = \frac{C}{(2-x)^2}$, $x \in [0,1]$. So $q^x(dy) = N(x)\delta_2(dy)$ is the unique nonnegative kernel achieving the minimum in $\mathcal{I}_c(\mu,\delta_2)$. Equivalently $\pi(dxdy) = N(x)\, dx \otimes \delta_2$ is the unique strong solution of the transport problem. It will follow from Theorem 2.2 below that all weak solutions are actually strong in this example, so π is also the unique weak solution.

1.2.3. *An example exhibiting both strong and weak solutions.* Let μ be the uniform distribution on $[0,1]$ and $\nu(dy) = 2y^2 \mathbf{1}_{[0,1]}(dy) + \frac{1}{3}\delta_0(dy)$. Consider the cost function

$$c(x,m) = \left| x - \int y\, dm \right|^p, \qquad x \in [0,1], m \in \mathcal{M}([0,1]),$$

with $0 < p$. We refer to Section 5.3 (in particular the proof of Theorem 5.3) for more insights about this type of costs and the construction of the weak solution below.

Let us first show that the transport problem (10) between μ and ν admits a strong solution. Consider the nonnegative kernel $\bar{q}^x(dy) = 2x\nu(dy)$. Since $\int x\, \mu(dx) = \int y\, \nu(dy) = \frac{1}{2}$, it is clear that $\mu q = \nu$ and that $\int y\, \bar{q}^x(dy) = x$. Therefore $\int c(x, \bar{q}^x)\, \mu(dx) = 0$, which shows that $\mathcal{I}_c(\mu,\nu) = 0$ and $\bar{q}$ is a strong solution.

Now let us construct a weak (but not strong) solution. Define $\eta(dx) = \sqrt{x}\mu(dx) + \frac{1}{3}\delta_0(dx)$ and let $\pi(dxdy) = \eta(dx)\delta_{\sqrt{x}}(dy)$. We claim that π is a weak solution of the transport problem (10) between μ and ν. First it is easy to check that the second marginal of π is ν, in other words that ν is the push-forward of η under the map $x \mapsto \sqrt{x}$. Let us now construct a minimizing sequence converging to π. Define π_ε, for $0 < \varepsilon < 1/2$ as the law of $((1-\varepsilon)X + \varepsilon U, \sqrt{X})$, where $X \sim \eta$ and $U \sim \mu$. A simple calculation shows that $\pi_\varepsilon(dxdy) = \mu(dx) q_\varepsilon^x(dy)$, where q is the nonnegative kernel defined by

$$q_\varepsilon^x(dy) := \frac{1}{3\varepsilon}\mathbf{1}_{a \leqslant \varepsilon}\delta_0(dy) + \frac{2y^2}{\varepsilon}\mathbf{1}_{\left[\sqrt{\max\left(\frac{x-\varepsilon}{1-\varepsilon},0\right)};\sqrt{\min\left(\frac{x}{1-\varepsilon},1\right)}\right]}(y)\, dy.$$

Therefore, for all $x \in [0,1]$,

$$b_\varepsilon(x) := \int y\, q_\varepsilon^x(dy) = \frac{1}{2\varepsilon}\left[\min\left(\frac{x}{1-\varepsilon},1\right)^2 - \max\left(\frac{x-\varepsilon}{1-\varepsilon},0\right)^2\right]$$
$$= \frac{1}{(1-\varepsilon)^2}\begin{cases} \frac{1}{2}\frac{x^2}{\varepsilon} & \text{if } 0 \leqslant x \leqslant \varepsilon \\ (x-\frac{\varepsilon}{2}) & \text{if } \varepsilon \leqslant x \leqslant 1-\varepsilon \\ \frac{1}{2}\frac{(1-x)}{\varepsilon}(1+x-2\varepsilon) & \text{if } 1-\varepsilon \leqslant x \leqslant 1 \end{cases}.$$

Thus, one sees that for all $x \in [0,1]$, $b_\varepsilon(x) \to x$ as $\varepsilon \to 0$ and that $\sup_{0<\varepsilon<1/2}\sup_{x\in[0,1]} b_\varepsilon(x) < +\infty$. So, applying the dominated convergence theorem yields

$$\int c(x, q_\varepsilon^x)\,\mu(dx) \to 0$$

as $\varepsilon \to 0$. Since $\pi^\varepsilon \to \pi$ in the weak sense, this shows that π is a weak solution (which is obviously not strong).

1.2.4. *A particular case of a one dimensional nonpositive conical cost function.* Consider the following cost function $c : [\alpha,\beta] \times \mathcal{M}([\gamma,\delta]) \to \mathbb{R}_-$ where $\alpha,\beta,\gamma,\delta \geqslant 0$.

$$c(x,m) = -x\left(\int y\,dm\right)^\eta, \qquad x \in [\alpha,\beta], m \in \mathcal{M}([\gamma,\delta]), \tag{12}$$

where $0 < \eta < 1$. This cost function is a particular case of the cost functions considered in [15] (in arbitrary dimensions). Note that c satisfies Assumption (LB). Indeed, by concavity of the function $y \mapsto y^\eta$ on $\mathbb{R}_+$, it holds

$$y^\eta \leqslant 1 + \eta(y-1), \qquad \forall y \geqslant 0.$$

Therefore,

$$c(x,m) \geqslant -x\left(1-\eta+\eta\int y\,dm\right) \geqslant -\beta\left(1-\eta+\eta\delta m([\gamma,\delta])\right) := r_0 + r_1 m([\gamma,\delta]).$$

The following result gives informations on strong solutions of the transport problem associated to the cost function c defined above.

Proposition 1.3. *Let $\mu \in \mathcal{P}([\alpha,\beta])$ and $\nu \in \mathcal{P}([\gamma,\delta])$; the transport problem (10) between μ and ν with respect to the cost function c defined by (12) admits strong solutions. For instance, denoting by $\bar{\mu}(dx) = \frac{1}{Z}x^{\frac{1}{1-\eta}}\,\mu(dx)$, where Z is a normalizing constant, then the coupling $\bar{\pi}$ given by*

$$\bar{\pi} = \bar{\mu} \otimes \nu$$

is a strong solution. More generally, $q \in \mathcal{Q}(\mu,\nu)$ is a nonnegative kernel solution if and only if there exists some constant $C > 0$ such that

$$\int y\, q^x(dy) = Cx^{\frac{1}{1-\eta}} \tag{13}$$

for μ almost all $x \in [\alpha,\beta]$. In particular, if μ has a positive density f on $[\alpha,\beta]$ and ν a positive density g on $[\gamma,\delta]$, and $T : [\alpha,\beta] \to [\gamma,\delta]$ is a continuously differentiable bijection such that for some constant $C > 0$ it holds

$$N(x)T(x) = Cx^{\frac{1}{1-\eta}}, \tag{14}$$

for Lebesgue almost all $x \in [\alpha,\beta]$, where N is the density (with respect to μ) defined by

$$N(x) = \frac{g(T(x))|T'(x)|}{f(x)}, \qquad \forall x \in [\alpha,\beta] \tag{15}$$

then $q^x(dy) = N(x)\delta_{T(x)}$, $x \in [\alpha,\beta]$, is a strong solution.

We will see in Theorem 5.4 below that for such cost function, all weak solutions are actually strong.

Proof. Let $\pi \in \Pi(\ll \mu, \nu)$, then $\frac{d\pi_1}{d\mu}(x) = \frac{d\pi_1}{d\bar{\mu}}(x)\frac{1}{Z}x^{\frac{1}{1-\eta}}$ and so

$$-I_c^\mu[\pi] = \int_\alpha^\beta x\left(\frac{d\pi_1}{d\mu}\right)^\eta \left(\int_\gamma^\delta y\,\pi^x(dy)\right)^\eta \mu(dx) = \frac{1}{Z^\eta}\int_\alpha^\beta x^{\frac{1}{1-\eta}}\left(\frac{d\pi_1}{d\bar{\mu}}\right)^\eta \left(\int_\gamma^\delta y\,\pi^x(dy)\right)^\eta \mu(dx)$$

$$= Z^{1-\eta}\int_\alpha^\beta \left(\frac{d\pi_1}{d\bar{\mu}}\right)^\eta \left(\int_\gamma^\delta y\,\pi^x(dy)\right)^\eta \bar{\mu}(dx) \leqslant Z^{1-\eta}\left(\int_\alpha^\beta \frac{d\pi_1}{d\bar{\mu}}(x)\int_\gamma^\delta y\,\pi^x(dy)\,\bar{\mu}(dx)\right)^\eta = Z^{1-\eta}\left(\int y\,\nu(dy)\right)^\eta,$$

where the inequality follows from the concavity of the function $u \mapsto u^\eta$. Note that if $\pi_1 = \bar{\mu}$ and $\pi^x(dy) = \nu(dy)$ for all x, there is equality. In other words, $\bar{\pi} = \bar{\mu} \otimes \nu$ is a strong solution of the transport problem between μ and ν. Moreover, according to the equality case in Jensen's inequality and the strict concavity of $u \mapsto u^\eta$, we see that there is equality above if and only if the function $x \mapsto \frac{d\pi_1}{d\bar{\mu}}(x)\int_\gamma^\delta y\,\pi^x(dy)$ is constant $\bar{\mu}$ almost surely. Writing $\pi(dxdy) = \mu(dx)q^x(dy)$, we see that this condition is equivalent to the existence of $C > 0$ such that (13) holds μ almost everywhere. Now, let us assume that μ has a positive density f on $[\alpha, \beta]$ and ν a positive density g on $[\gamma, \delta]$, and let us look for solutions of the form $q^x(dy) = N(x)\delta_{T(x)}$, where $x \mapsto T(x)$ is a continuously differentiable bijection from $[\alpha, \beta]$ to $[\gamma, \delta]$. First of all, if N satisfies (15), then for any bounded measurable function h on $[\gamma, \delta]$, it holds

$$\int_\alpha^\beta h(T(x))N(x)f(x)\,dx = \int_\alpha^\beta h(T(x))g(T(x))|T'(x)|\,dx = \int_\gamma^\delta h(y)g(y)\,dy,$$

by the change of variable formula, which shows that $\mu q = \nu$. Now, according to (14), it holds

$$\int y\,q^x(dy) = N(x)T(x) = Cx^{\frac{1}{1-\eta}},$$

for μ almost every $x \in [\alpha, \beta]$, which shows that q satisfies (13) and completes the proof. □

In the following result we consider the particular case where $\mu = \nu$ is the uniform measure on $[0, 1]$.

Corollary 1.1. *If μ and ν are both the uniform distribution on $[0, 1]$, the three following kernels are strong solutions of the problem:*

- Random sorting: $q_0^x(dy) = N_0(x)\mu(dy)$ *with* $N_0(x) = Cx^{a_0}$, *for all* $x \in [0, 1]$, $a_0 = 1/(1-\eta)$ *and* $C = (2-\eta)/(1-\eta)$ *is such that* $\int_0^1 N_0(x)\,\mu(dx) = 1$;
- Positive Assortative Matching: $q_1^x(dy) = N_1(x)\,\delta_{T_1(x)}$, *where for all* $x \in [0, 1]$,

$$T_1(x) = x^{a_1}, \quad N_1(x) = T_1'(x), \quad a_1 = \frac{2-\eta}{2(1-\eta)} = \frac{C}{2};$$

- Negative Assortative Matching: $q_2^x(dy) = N_2(x)\,\delta_{T_2(x)}$, *where for all* $x \in [0, 1]$,

$$T_2(x) = \sqrt{1 - x^{a_2}}, \quad N_2(x) = -T_2'(x), \quad a_2 = \frac{2-\eta}{1-\eta} = C.$$

Proof. The verification that q_1 and q_2 are strong solutions is left to the reader. □

2. Weak solutions as minimizers of an extended functional

As explained above, the difficulty in dealing with the minimization problem (11) is that the set $\Pi(\ll \mu, \cdot)$ is not closed in general, and so the optimal value of the problem can be reached at the boundary. In this section, we first identify the closure of $\Pi(\ll \mu, \nu)$, using a simple approximation technique from [28]. Then, we introduce an explicit functional $\bar{I}_c^\mu$ which is a lower semicontinuous extension of I_c^μ, and we introduce a condition (see (Approx) below) under which $\bar{I}_c^\mu$ coincides with the lower semicontinuous envelope of I_c^μ. When this condition is in force, we can interpret weak solutions as minimizers of $\bar{I}_c^\mu$ on the closure of $\Pi(\ll \mu, \nu)$. Finally, when Assumption (B) is satisfied, we will see that every weak solution is strong.

2.1. **Closure of $\Pi(\ll \mu, \nu)$.** Let us introduce a general mollifying approximation technique from [28, Theorem C.5], that will be very useful in the next paragraphs.

Lemma 2.1 (Lott-Villani [28])**.** *Let (S, d) be an arbitrary compact metric space and μ be a Borel probability measure on S. There exist a family of kernels $(K_n)_{n \geqslant 0}$ such that*

- *(i) For all $n \geqslant 0$, $K_n : S \times S \to \mathbb{R}_+$ is a continuous and symmetric function such that for all $x \in \mathrm{Supp}(\mu)$, $\int K_n(x,y)\mu(dy) = 1$.*
- *(ii) For all continuous function $f : \mathrm{Supp}(\mu) \to \mathbb{R}$, the functions $K_n f$, $n \geqslant 0$, defined by*

$$K_n f(y) := \int K_n(x,y) f(x)\, \mu(dx), \qquad y \in S, \tag{16}$$

  are continuous on $\mathrm{Supp}(\mu)$ and such that $K_n f \to f$ uniformly on $\mathrm{Supp}(\mu)$ as $n \to \infty$.
- *(iii) For all probability measure $\eta \in \mathcal{P}(S)$ such that $\eta(\mathrm{Supp}(\mu)) = 1$, the probability measures $K_n\eta$, $n \geqslant 0$, defined by*

$$K_n\eta(dy) := \int K_n(x,y)\, \eta(dx)\mu(dy)$$

  is such that $K_n\eta \to \eta$ as $n \to \infty$ for the weak convergence.

For a fixed $\mu \in \mathcal{P}(\mathcal{X})$, we will denote in what follows by $\Pi(\mathrm{Supp}(\mu), \nu)$ the set of probability measures π on $\mathcal{X} \times \mathcal{Y}$ such that $\pi_1(\mathrm{Supp}(\mu)) = 1$ and $\pi_2 = \nu$, where π_1 and π_2 denote respectively the marginals of π on $\mathcal{X}$ and $\mathcal{Y}$.

Lemma 2.2. *For any $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$, it holds*

$$\mathrm{cl}\, \Pi(\ll \mu, \nu) = \Pi(\mathrm{Supp}(\mu), \nu),$$

where $\mathrm{cl}\, \Pi(\ll \mu, \nu)$ denotes the closure of $\Pi(\ll \mu, \nu)$ for the weak topology. More precisely, for any $\pi \in \Pi(\mathrm{Supp}(\mu), \nu)$ with $\pi(dxdy) = \eta(dx)\pi^x(dy)$, the sequence $(\pi_n)_{n \geqslant 0}$ defined for all $n \geqslant 0$ by

$$\pi_n(dxdy) = \int K_n(x,z)\pi^z(dy)\eta(dz)\mu(dx),$$

where $(K_n)_{n \geqslant 0}$ is the sequence of kernels given by Lemma 2.1 (applied to $S = \mathcal{X}$ and μ) is such that $\pi_n \in \Pi(\ll \mu, \nu)$ for all $n \geqslant 0$ and $\pi_n \to \pi$ for the weak topology as $n \to \infty$.

Proof of Lemma 2.2. The inclusion $\subset$ is clear. Let us show the other inclusion. Let $\pi \in \Pi(\mathrm{Supp}(\mu), \nu)$ and set $\eta = \pi_1$. We claim that the first marginal of π_n is $K_n\eta$ and the second marginal is ν. Indeed, if $f : \mathcal{X} \to \mathbb{R}$ is a continuous function, then

$$\begin{aligned}\iint f(x)\,\pi_n(dxdy) &= \iiint f(x)K_n(z,x)\pi^z(dy)\mu(dx)\eta(dz)\\ &= \iint f(x)K_n(z,x)\mu(dx)\eta(dz)\\ &= \int f(x)(K_n\eta)(dx)\end{aligned}$$

and if $g : \mathcal{Y} \to \mathbb{R}$ is a continuous function, then

$$\begin{aligned}\iint g(y)\,\pi_n(dxdy) &= \iiint g(y)K_n(z,x)\pi^z(dy)\mu(dx)\eta(dz)\\ &= \iint g(y)\pi^z(dy)\eta(dz)\\ &= \int g(y)\nu(dy).\end{aligned}$$

If $f : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ is a continuous function, then denoting by f_y the function $x \mapsto f(x,y)$, it follows from Item (ii) of Lemma 2.1 that

$$\iint f(x,y)\,\pi_n(dxdy) = \iiint f(x,y)K_n(z,x)\mu(dx)\pi^z(dy)\eta(dz) = \iint K_n f_y(z)\,\pi^z(dy)\eta(dz) \to \iint f(y,z)\pi(dydz),$$

as $n\to\infty$. In other words $\pi_n \to \pi$ in the weak topology. Also, since $K_n\eta \ll \mu$, π_n belongs to $\Pi(\ll\mu,\nu)$ which completes the proof. □

2.1.1. *Lower semicontinuous extensions of I_c^μ.* For any fixed $\mu\in\mathcal{P}(\mathcal{X})$, consider the functional

$$\bar{I}_c^\mu : \mathcal{P}(\mathcal{X}\times\mathcal{Y}) \to \mathbb{R}\cup\{+\infty\}$$

defined by

$$\bar{I}_c^\mu[\pi] = \int c\left(x, \frac{d\pi_1^{ac}}{d\mu}(x)\pi^x(dy)\right)\,\mu(dx) + \int c'_\infty\left(x,\pi^x\right)\,\pi_1^s(dx), \qquad \forall \pi\in\mathcal{P}(\mathcal{X}\times\mathcal{Y}),$$

where $\pi_1 = \pi_1^{ac} + \pi_1^s$ is the decomposition of π_1 into an absolutely continuous part and a singular part with respect to μ and

$$c'_\infty(x,m) = \lim_{\lambda\to\infty}\frac{c(x,\lambda m)}{\lambda}, \qquad x\in\mathcal{X}, m\in\mathcal{M}(\mathcal{Y})$$

is the recession function of $c(x,\cdot)$. Note that this limit is always well defined since, by convexity of $c(x,\cdot)$, the function $\lambda\mapsto\frac{c(x,\lambda m)-c(x,0)}{\lambda}$ is non-decreasing on $(0,\infty)$.

The following proposition shows that, for a fixed μ and under Assumption (A), the functional $\bar{I}_c^\mu$ is a lower semicontinuous extension of I_c^μ.

Proposition 2.1. *Under Assumption* (A)*, the function $\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{X}\times\mathcal{Y}) : (\mu,\pi)\mapsto \bar{I}_c^\mu[\pi]$ is lower semicontinuous and such that $\bar{I}_c^\mu = I_c^\mu$ on $\Pi(\ll\mu,\cdot)$.*

The proof of Proposition 2.1 (which is adapted from [3]) is postponed to Section A.1 of Appendix.

For a fixed $\mu\in\mathcal{P}(\mathcal{X})$, let us now introduce the lower semicontinuous envelope of I_c^μ, denoted $\tilde{I}_c^\mu$ and defined as follows : for all $\pi\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$

$$\tilde{I}_c^\mu[\pi] = \sup_{V\in\mathcal{V}(\pi)}\ \inf_{\gamma\in V\cap\Pi(\ll\mu,\cdot)} I_c^\mu[\gamma],$$

where $\mathcal{V}(\pi)$ denotes the class of all open neighborhoods of π. By convention $\inf\varnothing = +\infty$, so in particular, $\tilde{I}_c^\mu = +\infty$ outside $\mathrm{cl}\,\Pi(\ll\mu,\cdot) = \Pi(\mathrm{Supp}(\mu),\cdot)$.

At this level of generality, it is not clear whether $\bar{I}_c^\mu$ and $\tilde{I}_c^\mu$ always coincide on $\Pi(\mathrm{Supp}(\mu),\,\cdot\,)$. The following proposition gives a necessary and sufficient condition for that.

Proposition 2.2. *For a fixed $\mu \in \mathcal{P}(\mathcal{X})$ and under Assumption (A), it holds*

(i) for all $\pi \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$, $\bar{I}_c^\mu[\pi] \leqslant \tilde{I}_c^\mu[\pi]$,
(ii) for all $\pi \in \Pi(\ll \mu,\,\cdot\,)$, $I_c^\mu[\pi] = \bar{I}_c^\mu[\pi] = \tilde{I}_c^\mu[\pi]$,
(iii) the functionals $\bar{I}_c^\mu$ and $\tilde{I}_c^\mu$ coincide on $\Pi(\mathrm{Supp}(\mu),\,\cdot\,)$ if and only if for all $\pi \in \Pi(\mathrm{Supp}(\mu),\,\cdot\,)$ there exists a sequence $\pi_n \in \Pi(\ll \mu,\,\cdot\,)$ such that $I_c^\mu[\pi_n] \to \bar{I}_c^\mu[\pi]$.

Proof. Since $\bar{I}_c^\mu$ is lower semicontinuous, for all $\pi \in \mathcal{P}(\mathcal{X}\times\mathcal{Y})$ it holds

$$\bar{I}_c^\mu[\pi] = \sup_{V\in\mathcal{V}(\pi)} \inf_{\gamma\in V} \bar{I}_c^\mu[\gamma] \leqslant \sup_{V\in\mathcal{V}(\pi)} \inf_{\gamma\in V\cap\Pi(\ll\mu,\,\cdot\,)} \bar{I}_c^\mu[\gamma] = \sup_{V\in\mathcal{V}(\pi)} \inf_{\gamma\in V\cap\Pi(\ll\mu,\,\cdot\,)} I_c^\mu[\gamma] = \tilde{I}_c^\mu[\pi],$$

and so $\bar{I}_c^\mu \leqslant \tilde{I}_c^\mu$, which proves (i). On the other hand, if $\pi \in \Pi(\ll \mu,\,\cdot\,)$, then

$$\inf_{\gamma\in V\cap\Pi(\ll\mu,\,\cdot\,)} I_c^\mu[\gamma] \leqslant I_c^\mu[\pi] = \bar{I}_c^\mu[\pi]$$

and so, optimizing over $V \in \mathcal{V}(\pi)$, $\tilde{I}_c^\mu[\pi] \leqslant \bar{I}_c^\mu[\pi]$ which proves (ii). Let us prove (iii). Suppose that $\pi \in \Pi(\mathrm{Supp}(\mu),\,\cdot\,)$ is such that there exists a sequence $\pi_n \in \Pi(\ll \mu,\,\cdot\,)$ for which $I_c^\mu[\pi_n] \to \bar{I}_c^\mu[\pi]$. Then, since $\tilde{I}_c^\mu$ is lower semicontinuous, it holds

$$\tilde{I}_c^\mu[\pi] \leqslant \liminf_{n\to\infty} \tilde{I}_c^\mu[\pi_n] = \liminf_{n\to\infty} I_c^\mu[\pi_n] = \bar{I}_c^\mu[\pi].$$

Since the inequality $\bar{I}_c^\mu[\pi] \leqslant \tilde{I}_c^\mu[\pi]$ is always true, there is in fact equality. Conversely, suppose that $\tilde{I}_c^\mu = \bar{I}_c^\mu$ on $\Pi(\mathrm{Supp}(\mu,\,\cdot\,)$. If $\pi \in \Pi(\mathrm{Supp}(\mu),\,\cdot\,)$, then according to Lemma 2.2, $\pi \in \mathrm{cl}\,\Pi(\ll \mu,\,\cdot\,)$. Therefore, for any open neighborhood V of π, the set $V\cap\Pi(\ll \mu,\,\cdot\,)$ is non-empty. Now, it easily follows from the definition of $\tilde{I}_c^\mu$, that there exists some sequence $\pi_n \in \Pi(\ll \mu,\,\cdot\,)$ such that $I_c^\mu[\pi_n] \to \tilde{I}_c^\mu[\pi]$ and so $I_c^\mu[\pi_n] \to \bar{I}_c^\mu[\pi]$. □

For a fixed $\mu \in \mathcal{P}(\mathcal{X})$, let us introduce the following variants of problem (10): for $\nu \in \mathcal{P}(\mathcal{Y})$,

$$\bar{\mathcal{I}}_c(\mu,\nu) = \inf_{\pi\in\Pi(\mathrm{Supp}(\mu),\nu)} \bar{I}_c^\mu[\pi] \tag{17}$$

and

$$\tilde{\mathcal{I}}_c(\mu,\nu) = \inf_{\pi\in\Pi(\mathrm{Supp}(\mu),\nu)} \tilde{I}_c^\mu[\pi]. \tag{18}$$

Unlike transport problem (10), the transport problems (17) and (18) always admit solutions.

Lemma 2.3. *Under Assumption (A), for any $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ such that $\bar{\mathcal{I}}_c(\mu,\nu) < \infty$, there exists $\pi \in \Pi(\mathrm{Supp}\,(\mu),\nu)$ such that $\bar{I}_c^\mu[\pi] = \bar{\mathcal{I}}_c(\mu,\nu)$. The same is true for the transport problem (18).*

Proof. The functional $\bar{I}_c^\mu$ is lower semicontinuous on the compact set $\Pi(\mathrm{Supp}(\mu),\nu)$ so it attains its lower bound. □

Finally, the following result will be very useful in Section 3 dealing with duality.

Proposition 2.3. *Under Assumption (A), the functional $\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y}) \to \mathbb{R}\cup\{+\infty\} : (\mu,\nu) \mapsto \bar{\mathcal{I}}_c(\mu,\nu)$ is lower semicontinuous at any point (μ,ν) with* $\mathrm{Supp}\,(\mu) = \mathcal{X}$.

Proof. Let $(\mu,\nu) \in \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$ be such that $\mathrm{Supp}\,(\mu) = \mathcal{X}$ and consider $(\mu_n,\nu_n) \in \mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})$ a sequence converging to (μ,ν). According to Lemma 2.3, for all $n \geqslant 0$ there exists $\pi_n \in \Pi(\mathrm{Supp}(\mu_n),\nu_n)$ such that

$$\bar{I}_c^{\mu_n}[\pi_n] = \bar{\mathcal{I}}_c(\mu_n,\nu_n).$$

Let $\ell = \liminf_{n\to\infty} \bar{I}_c^{\mu_n}[\pi_n]$. Extracting a subsequence if necessary, one can assume without loss of generality that $\bar{I}_c^{\mu_n}[\pi_n] \to \ell$ as $n \to \infty$. Since $\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ is compact, the sequence π_n admits a converging subsequence, that we will denote again by π_n. Let $\bar{\pi}$ be the limit of π_n. Since $\nu_n \to \nu$, $\bar{\pi} \in \Pi(\mathcal{X},\nu) = \Pi(\mathrm{Supp}\,(\mu),\nu)$. Then, by semicontinuity of $\bar{I}_c^{\cdot}[\,\cdot\,]$, it holds

$$\ell = \lim_{n\to\infty} \bar{I}_c^{\mu_n}[\pi_n] \geqslant \bar{I}_c^{\mu}[\bar{\pi}] \geqslant \inf_{\pi\in\Pi(\mathrm{Supp}\,(\mu),\nu)} \bar{I}_c^{\mu}[\pi] = \bar{\mathcal{I}}_c(\mu,\nu).$$

□

2.2. **Weak solutions as minimizers of $\bar{I}_c^{\mu}$.** The following inequality is always true

$$\bar{\mathcal{I}}_c(\mu,\nu) \leqslant \widetilde{\mathcal{I}}_c(\mu,\nu) \leqslant \mathcal{I}_c(\mu,\nu). \tag{19}$$

Indeed, the first inequality comes from the fact that $\bar{I}_c^{\mu} \leqslant \tilde{I}_c^{\mu}$ (Item (i) of Proposition 2.2) and the second from the fact that $\tilde{I}_c^{\mu} = I_c^{\mu}$ on $\Pi(\ll \mu,\nu) \subset \Pi(\mathrm{Supp}(\mu),\nu)$. It is not clear if there is always equality in (19). The following result gives a sufficient condition.

We will say that c satisfies Assumption (Approx) if for all $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$,

(Approx) $\quad \forall \pi \in \Pi(\mathrm{Supp}(\mu),\nu)$ there exists a sequence $\pi_n \in \Pi(\ll \mu,\nu)$ such that $I_c^{\mu}[\pi_n] \to \bar{I}_c^{\mu}[\pi]$.

Of course, when $\pi \in \Pi(\ll \mu,\nu)$, one can choose the constant sequence $\pi_n = \pi$, $n \geqslant 0$. Only the case $\pi \in \Pi(\mathrm{Supp}(\mu),\nu)\backslash\Pi(\ll \mu,\nu)$ is non trivial in the above condition. Note that this condition is trivially satisfied when $\mathcal{X}$ is finite. We will see below more general sufficient conditions for (Approx).

Theorem 2.1. *Let $c : \mathcal{X}\times\mathcal{M}(\mathcal{Y}) \to \mathbb{R}$ be a cost function satisfying condition (A) and (Approx) and $\mu \in \mathcal{P}(\mathcal{X})$.*

- (i) *For any $\pi \in \Pi(\mathrm{Supp}(\mu),\,\cdot\,)$, it holds $\bar{I}_c^{\mu}[\pi] = \tilde{I}_c^{\mu}[\pi]$.*
- (ii) *For any $\nu \in \mathcal{P}(\mathcal{Y})$, it holds $\mathcal{I}_c(\mu,\nu) = \widetilde{\mathcal{I}}_c(\mu,\nu) = \bar{\mathcal{I}}_c(\mu,\nu)$.*
- (iii) *Let $\nu \in \mathcal{P}(\mathcal{Y})$ be such that $\mathcal{I}_c(\mu,\nu) < +\infty$. A coupling $\pi \in \Pi(\mathrm{Supp}(\mu),\nu)$ is a weak solution of the transport problem (10) if and only if π minimizes $\bar{I}_c^{\mu}$ on $\Pi(\mathrm{Supp}(\mu),\nu)$.*

Proof. Item (i) follows from Proposition 2.2 (Item (iii)) and Assumption (Approx). Let us show Item (ii). Let $\pi \in \Pi(\mathrm{Supp}(\mu),\nu)$ and consider a sequence $\pi_n \in \Pi(\ll \mu,\nu)$ such that $I_c^{\mu}[\pi_n] \to \bar{I}_c^{\mu}[\pi]$. For all n, it holds $\mathcal{I}_c(\mu,\nu) \leqslant I_c^{\mu}[\pi_n]$, and so letting $n \to \infty$ gives $\mathcal{I}_c(\mu,\nu) \leqslant \bar{I}_c^{\mu}[\pi]$. Optimizing over all $\pi \in \Pi(\mathrm{Supp}(\mu),\nu)$ yields to $\mathcal{I}_c(\mu,\nu) \leqslant \bar{\mathcal{I}}_c(\mu,\nu)$, which together with (19) proves the claim. Let us finally show Item (iii). Since $\bar{\mathcal{I}}_c(\mu,\,\cdot\,) = \mathcal{I}_c(\mu,\,\cdot\,)$, it follows that any minimizer of $\bar{I}_c^{\mu}$ on $\Pi(\mathrm{Supp}(\mu),\nu) = \mathrm{cl}\,\Pi(\ll \mu,\nu)$ is a weak solution. Conversely, note that if π_n is a sequence of $\Pi(\ll \mu,\nu)$ converging to some π and such that $I_c^{\mu}[\pi_n] \to \mathcal{I}_c(\mu,\nu)$, then by lower semicontinuity of $\bar{I}_c^{\mu}$, it holds $\bar{I}_c^{\mu}[\pi] \leqslant \mathcal{I}_c(\mu,\nu) = \bar{\mathcal{I}}_c(\mu,\nu)$, and so $\pi \in \Pi(\mathrm{Supp}(\mu),\nu)$ is a minimizer of $\bar{I}_c^{\mu}$ on $\Pi(\mathrm{Supp}(\mu),\nu)$, which completes the proof. □

Let us now give some concrete conditions on c ensuring (Approx). We will say that a cost function $c : \mathcal{X}\times\mathcal{M}(\mathcal{Y}) \to \mathbb{R}$ satisfies Assumption (C) if

$$\text{(C)} \qquad \begin{cases} - \ \text{for all } m \in \mathcal{M}(\mathcal{Y}),\ \text{the functions } c(\,\cdot\,,m) \text{ and } c'_\infty(\,\cdot\,,m) \text{ are continuous on } \mathcal{X}, \\ \text{and} \\ - \ \text{there exists } a \geqslant 0 \text{ such that } c'_\infty(x,p) \leqslant a \text{ for all } x \in \mathcal{X} \text{ and } p \in \mathcal{P}(\mathcal{Y}). \end{cases}$$

For instance, the cost function introduced in Section 1.2.4 :

$$c(x,m) = -x\left(\int y\,dm\right)^{\eta}, \qquad x \in [\alpha,\beta], m \in \mathcal{M}([\gamma,\delta]),$$

where $\alpha, \beta, \gamma, \delta \geqslant 0$, $\eta \in (0,1)$, is such that

$$c'_\infty(x,p) = 0, \qquad x \in [\alpha,\beta], p \in \mathcal{P}([\gamma,\delta]),$$

and so c satisfies Assumption (C).

Lemma 2.4. *If $c : \mathcal{X} \times \mathcal{M}(\mathcal{Y}) \to \mathbb{R}$ satisfies Assumptions (A) and (C), then it satisfies Assumption (Approx). More precisely, for any $\mu \in \mathcal{P}(\mathcal{X})$ and $\pi \in \Pi(\mathrm{Supp}(\mu),\nu)$, the sequence $\pi_n \in \Pi(\ll \mu,\nu)$, $n \geqslant 0$, defined in Lemma 2.2 is such that $I_c^\mu(\pi_n) \to \bar{I}_c^\mu(\pi)$.*

The proof of Lemma 2.4 is postponed to Section A.2 of Appendix.

2.3. **A criterion for the existence of strong solutions.** Recall Assumption (B) given in the introduction, which can be recast as follows :

$$c'_\infty(x,m) = +\infty, \qquad \forall m \in \mathcal{M}(\mathcal{Y})\backslash\{0\}, \qquad \forall x \in \mathcal{X}.$$

Under Assumption (B), one gets $\bar{I}_c^\mu[\pi] = I_c^\mu[\pi]$, if $\pi \in \Pi(\ll \mu, \cdot)$ and $+\infty$ otherwise.

Lemma 2.5. *If $c : \mathcal{X} \times \mathcal{M}(\mathcal{Y}) \to \mathbb{R}$ is a cost function satisfying Assumptions (A) and (B), then it satisfies (Approx).*

Proof. If $\pi_n \in \Pi(\ll \mu,\nu)$ is any sequence converging to $\pi \in \Pi(\mathrm{Supp}(\mu),\nu)\backslash\Pi(\ll \mu,\nu)$ (such sequences always exist according to Lemma 2.2), then since $\bar{I}_c^\mu$ is lower semicontinuous, one gets

$$\liminf_{n\to\infty} I_c^\mu[\pi_n] = \liminf_{n\to\infty} \bar{I}_c^\mu[\pi_n] \geqslant \bar{I}_c^\mu[\pi] = +\infty$$

and so $I_c^\mu[\pi_n] \to \bar{I}_c^\mu[\pi]$. □

The following result shows in particular that strong solutions always exist under Assumptions (A) and (B).

Theorem 2.2. *Let $c : \mathcal{X} \times \mathcal{M}(\mathcal{Y}) \to \mathbb{R}$ be a cost function satisfying Assumptions (A) and (B). If $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$ are such that $\mathcal{I}_c(\mu,\nu) < +\infty$, then any weak solution of the transport problem (10) is a strong solution.*

Proof. According to Lemma 2.5 and Theorem 2.1, if π is a weak solution, then

$$\bar{I}_c^\mu[\pi] = \bar{\mathcal{I}}_c(\mu,\nu) = \mathcal{I}_c(\mu,\nu) < \infty.$$

Therefore, $\bar{I}_c^\mu[\pi] < +\infty$ and so $\pi \in \Pi(\ll \mu,\nu)$ and $\bar{I}_c^\mu[\pi] = I_c^\mu[\pi] = \mathcal{I}_c(\mu,\nu)$, which shows that π is a strong solution. □

Note that condition (B) applies for instance if there exists $\phi : \mathbb{R}_+ \to \mathbb{R}$ a function such that $\phi(u)/u \to +\infty$, when $u \to +\infty$, such that

$$c(x,m) \geqslant \phi(m(\mathcal{Y})), \qquad \forall x \in \mathcal{X}, \forall m \in \mathcal{M}(\mathcal{Y}). \tag{20}$$

If $\mathcal{X}, \mathcal{Y} \subset \mathbb{R}^d$ and the convex hull of $\mathcal{Y}$ does not contain 0, this assumption is for instance satisfied by the following conical cost functions

$$c(x,m) = \left\| x - \int y\, dm \right\|^p, \qquad x \in \mathcal{X}, m \in \mathcal{M}(\mathcal{Y}), \tag{21}$$

where $\|\cdot\|$ is an arbitrary norm on $\mathbb{R}^d$ and $p > 1$.

Remark 2.1. *Let us briefly indicate another possible method for proving existence of strong solutions when c satisfies (20). Let $\nu \in \mathcal{P}(\mathcal{Y})$ be such that $\mathcal{I}_c(\mu,\nu) < +\infty$ and assume $(\pi_n)_{n\geqslant 0}$ is a sequence in $\Pi(\ll \mu,\nu)$ such that $I_c^\mu[\pi_n] \to \mathcal{I}_c(\mu,\nu)$ and $\pi_n \to \pi$. Then it follows from (20) that $\sup_{n\in\mathbb{N}} \int \phi(N_n)\,d\mu < +\infty$, denoting by N_n the density of the first marginal of π_n with respect to μ. Therefore, the sequence $(N_n)_{n\geqslant 0}$ is uniformly integrable and so, according to the Dunford-Pettis theorem, it admits a converging subsequence for the topology $\sigma(L^1(\mu), L^\infty(\mu))$. From this follows easily that $\pi \in \Pi(\ll \mu,\nu)$ and is therefore a strong solution.*

3. Dual formulations

In this section, we establish a Kantorovich type dual formula for the transport problem (3). The derivation of these dual forms will make use of the following abstract Fenchel-Moreau biconjugation theorem (see e.g [36, Theorem 2.3.4]).

Theorem 3.1. *Let E be a Hausdorff locally convex topological vector space and E' its topological dual space. If $F : E \to (-\infty,\infty]$ is a convex function such that $\mathrm{dom}\,(F) = \{x \in E : F(x) < \infty\} \neq \varnothing$, then for any $x \in \mathrm{dom}\,(F)$ where F is lower semicontinuous, it holds*

$$F(x) = \sup_{\ell\in E'} \{\ell(x) - F^*(\ell)\}$$

where

$$F^*(\ell) = \sup_{x\in E} \{\ell(x) - F(x)\}, \qquad \ell \in E'.$$

In what follows, we will apply Theorem 3.1 in the following setting: $E = \mathcal{M}_s(\mathcal{X}) \times \mathcal{M}_s(\mathcal{Y})$ equipped with the product weak topology whose topological dual is $E' = \mathcal{C}_b(\mathcal{X}) \times \mathcal{C}_b(\mathcal{Y})$ and $F : \mathcal{M}_s(\mathcal{X}) \times \mathcal{M}_s(\mathcal{Y}) \to \mathbb{R} \cup \{+\infty\}$ defined as follows

$$F(\alpha,\beta) = \begin{cases} \mathcal{I}_c\left(\frac{\alpha}{\alpha(\mathcal{X})}, \frac{\beta}{\beta(\mathcal{Y})}\right)\alpha(\mathcal{X}) & \text{if } \alpha,\beta \geqslant 0 \text{ and } \alpha(\mathcal{X}) = \beta(\mathcal{Y}) > 0 \\ 0 & \text{if } \alpha,\beta \geqslant 0 \text{ and } \alpha(\mathcal{X}) = \beta(\mathcal{Y}) = 0 \\ +\infty & \text{otherwise} \end{cases}, \tag{22}$$

for all $\alpha \in \mathcal{M}_s(\mathcal{X})$ and $\beta \in \mathcal{M}_s(\mathcal{Y})$.

Lemma 3.1. *The functional F is convex on $\mathcal{M}_s(\mathcal{X}) \times \mathcal{M}_s(\mathcal{Y})$. Moreover, under Assumptions (A) and (Approx), the functional F is lower semicontinuous at any point (α,β) such that $\alpha,\beta \geqslant 0$, $\alpha(\mathcal{X}) = \beta(\mathcal{Y})$ and $\mathrm{Supp}\,(\alpha) = \mathcal{X}$.*

Proof. The first statement easily follows from Proposition 1.1. Let α_n, β_n be sequences converging respectively to finite nonnegative measures α, β such that $\alpha(\mathcal{X}) = \beta(\mathcal{Y})$ and $\mathrm{Supp}\,(\alpha) = \mathcal{X}$ and let us show that $\liminf_{n\to\infty} F(\alpha_n,\beta_n) \geqslant F(\alpha,\beta)$. Dropping terms if necessary, one can assume without loss of generality that $\alpha_n(\mathcal{X}) = \beta_n(\mathcal{Y})$ for all n. As α has full support, $\alpha(\mathcal{X}) > 0$ and since $\alpha_n(\mathcal{X}) \to \alpha(\mathcal{X})$, it follows that $\alpha_n(\mathcal{X}) > 0$ for all n large enough. Under Assumptions (A) and (Approx), Theorem 2.1 gives that $\mathcal{I}_c = \overline{\mathcal{I}}_c$. Since $\alpha_n/\alpha_n(\mathcal{X}) \to \alpha/\alpha(\mathcal{X})$ (which has full support) and $\beta_n/\beta_n(\mathcal{Y}) \to \beta/\beta(\mathcal{Y})$, it follows from Proposition 2.3 that

$$\liminf_{n\to\infty} \mathcal{I}_c\left(\frac{\alpha_n}{\alpha_n(\mathcal{X})}, \frac{\beta_n}{\beta_n(\mathcal{Y})}\right) \geqslant \mathcal{I}_c\left(\frac{\alpha}{\alpha(\mathcal{X})}, \frac{\beta}{\beta(\mathcal{Y})}\right)$$

and since $\alpha_n(\mathcal{X}) = \beta_n(\mathcal{Y}) \to \alpha(\mathcal{Y}) = \beta(\mathcal{Y})$ this proves the claim. □

Theorem 3.2. *Under Assumptions (A) and (Approx), it holds*

$$\mathcal{I}_c(\mu,\nu) = \sup_{f\in\mathcal{C}_b(\mathcal{Y})} \left\{ \int K_c f(x)\,\mu(dx) - \int f(y)\,\nu(dy) \right\}, \qquad \forall \nu \in \mathcal{P}(\mathcal{Y}) \tag{23}$$

where

$$K_c f(x) = \inf_{m \in \mathcal{M}(\mathcal{Y})} \left\{ \int f\, dm + c(x,m) \right\}, \qquad x \in \mathcal{X}.$$

In particular, (23) *holds whenever* c *satisfies Assumption* (A) *and Assumption* (B) *or* (C).

Remark 3.1. *It would be very interesting to obtain general sufficient conditions for dual attainment in Theorem 3.2. This could lead to cyclical monotonicity criterium characterizing optimality of transport plans, in the spirit of the* C*-monotonicity criterium obtained for WOT* [4, 7].

The proof below is adapted from the proof of [2, Theorem 4.2].

Proof. Let $\mu \in \mathcal{P}(\mathcal{X})$, $\nu \in \mathcal{P}(\mathcal{Y})$ and assume that μ has full support. According to Lemma 3.1, the function F defined by (22) is convex on $\mathcal{M}_s(\mathcal{X}) \times \mathcal{M}_s(\mathcal{Y})$ and lower semicontinuous at (μ,ν). Therefore, according to Theorem 3.1, it holds

$$\mathcal{I}_c(\mu,\nu) = F(\mu,\nu) = \sup_{(\varphi,\psi)\in \mathcal{C}_b(\mathcal{X})\times\mathcal{C}_b(\mathcal{Y})} \left\{ \int \varphi\, d\mu + \int \psi\, d\nu - F^*(\varphi,\psi) \right\},$$

with, for all $(\varphi,\psi) \in \mathcal{C}_b(\mathcal{X}) \times \mathcal{C}_b(\mathcal{Y})$,

$$\begin{aligned} F^*(\varphi,\psi) &= \sup_{(\alpha,\beta)\in\mathcal{M}_s(\mathcal{X})\times\mathcal{M}_s(\mathcal{Y})} \left\{ \int \varphi\, d\alpha + \int \psi\, d\beta - F(\alpha,\beta) \right\} \\ &= \sup_{(\alpha,\beta)\in\mathcal{P}(\mathcal{X})\times\mathcal{P}(\mathcal{Y})} \sup_{\lambda \geqslant 0} \left\{ \lambda \int \varphi\, d\alpha + \lambda \int \psi\, d\beta - \lambda \mathcal{I}_c(\alpha,\beta) \right\} \\ &= \chi_{\mathcal{K}_c}(\varphi,\psi), \end{aligned}$$

where

$$\mathcal{K}_c = \left\{ (\varphi,\psi) \in \mathcal{C}_b(\mathcal{X}) \times \mathcal{C}_b(\mathcal{Y}) : \int \varphi\, d\alpha + \int \psi\, d\beta \leqslant \mathcal{I}_c(\alpha,\beta), \forall \alpha \in \mathcal{P}(\mathcal{X}), \forall \beta \in \mathcal{P}(\mathcal{Y}) \right\}$$

and $\chi_{\mathcal{K}_\alpha}(\varphi,\psi) = 0$ if $(\varphi,\psi) \in \mathcal{K}_c$ and $+\infty$ otherwise. Thus we get

$$\mathcal{I}_c(\mu,\nu) = F(\mu,\nu) = \sup_{(\varphi,\psi)\in\mathcal{K}_c} \left\{ \int \varphi\, d\mu + \int \psi\, d\nu \right\},$$

Now, observe that if $(\varphi,\psi) \in \mathcal{K}_c$, then (choosing $\alpha = \delta_x$, with $x \in \mathcal{X}$) it holds

$$\varphi(x) \leqslant \inf_{\beta\in\mathcal{P}(\mathcal{Y})} \left\{ -\int \psi\, d\beta + \mathcal{I}_c(\delta_x,\beta) \right\} \leqslant \inf_{\beta\in\mathcal{P}(\mathcal{Y})} \left\{ -\int \psi\, d\beta + c(x,\beta) \right\} = K_c(-\psi)(x),$$

where we used that $\mathcal{I}_c(\delta_x,\beta) = c(x,\beta)$. Thus, it holds

$$\mathcal{I}_c(\mu,\nu) \leqslant \sup_{f\in\mathcal{C}_b(\mathcal{Y})} \left\{ \int K_c f(x)\, \mu(dx) - \int f(y)\, \nu(dy) \right\}.$$

The converse inequality is always true. Indeed, if $\pi(dxdy) = N(x)\mu(dx)\pi^x(dy) \in \Pi(\ll \mu,\nu)$, then

$$\int K_c f(x)\, \mu(dx) \leqslant \int \left(\int f(y) d(N(x)\pi^x)(dy) + c(x, N(x)\pi^x) \right) \mu(dx) = \int f\, d\nu + I_c^\mu[\pi].$$

Formula (23) is thus proved when μ has full support. When μ does not have full support, then letting $\tilde{\mathcal{X}} = \mathrm{Supp}(\mu)$ and applying the preceding reasoning in the space $\mathcal{M}_s(\tilde{\mathcal{X}}) \times \mathcal{M}_s(\mathcal{Y})$ gives the desired duality formula. □

4. Monotonicity properties and uniqueness of primal solutions

In this section, we consider cost functions of the following form

$$c(x,m) = G\left(\int F(x,y)\,m(dy)\right), \qquad x \in \mathcal{X}, m \in \mathcal{M}(\mathcal{Y}), \tag{24}$$

where $F : \mathcal{X} \times \mathcal{Y} \to (0,+\infty)$ is a continuous function and $G : \mathbb{R}^+ \to \mathbb{R}$ is a convex function, assumed to be differentiable on $(0,+\infty)$ and we denote by $G'(0) = \lim_{x\to 0^+} G'(x) \in \mathbb{R} \cup \{-\infty\}$ and $G'(+\infty) = \lim_{x\to+\infty} G'(x) \in \mathbb{R} \cup \{+\infty\}$. We will establish below that the dual problem admits a solution (Theorem 4.1) and then use this dual optimizer to get informations on the support of primal solutions (Proposition 4.1). Finally, we will consider the particular case when $\mathcal{X}$ and $\mathcal{Y}$ are subsets of $\mathbb{R}$ and prove uniqueness of primal solutions under suitable assumptions on F, G and μ (Theorem 4.2).

First let us check that c satisfies the assumptions introduced in the preceding sections. Writing G as a countable supremum of affine functions, one easily sees that c satisfies Assumption (A) and in particular (LB). Thus $\mathcal{I}_c(\mu,\nu)$ makes sense for any $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$. Note also that for any $m \in \mathcal{M}(\mathcal{Y})\backslash\{0\}$ and $x \in \mathcal{X}$

$$c'_\infty(x,m) = G'(+\infty).$$

So c satisfies Assumption (B) if $G'(+\infty) = +\infty$ and Assumption (C) otherwise. Therefore, Theorem 3.2 applies and it is easily seen that

$$\mathcal{I}_c(\mu,\nu) = \sup_{f\in\mathcal{L}^1(\nu)} \left\{\int K_c f(x)\,\mu(dx) - \int f(y)\,\nu(dy)\right\}, \qquad \forall \mu \in \mathcal{P}(\mathcal{X}), \forall \nu \in \mathcal{P}(\mathcal{Y}), \tag{25}$$

with

$$\mathcal{L}^1(\nu) = \left\{f : \mathcal{Y} \to \mathbb{R} : f \text{ measurable and } \int |f|\,d\nu < +\infty\right\}$$

and

$$K_c f(x) = \inf_{m\in\mathcal{M}(\mathcal{Y}) \text{ s.t } f\in L^1(m)} \left\{\int f\,dm + c(x,m)\right\}, \qquad x \in \mathcal{X}, f \in \mathcal{L}^1(\nu). \tag{26}$$

Note that $K_c f$ is upper semicontinuous on $\mathcal{X}$ as an infimum of continuous functions and that $K_c f(x) \leqslant G(0)$ for all $x \in \mathcal{X}$ and thus $\int K_c f(x)\,\mu(dx)$ always makes sense.

The following result establishes dual attainment.

Theorem 4.1. *If the function G in (24) is such that $G'(0) > -\infty$, then for every $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$, there exists a bounded function $\bar{f}$ on $\mathcal{Y}$ such that*

$$\mathcal{I}_c(\mu,\nu) = \int K_c\bar{f}(x)\,\mu(dx) - \int \bar{f}(y)\,\nu(dy).$$

The same conclusion holds if $G'(0) = -\infty$ and $\mathcal{Y}$ is a finite set. Moreover, if G is non-decreasing (resp. non-increasing) on $\mathbb{R}^+$, then $\bar{f}$ can be chosen nonpositive (resp. nonnegative).

Proof. Let us show that the supremum in (25) can be restricted to the class of measurable functions f such that $f \geqslant a$, with $a = \mathcal{I}_c(\mu,\nu) - \sup_{u\in\mathcal{X},v\in\mathcal{Y}} G(2F(u,v)) - 1$. Note that if f is not bounded from below, then $K_c(f)(x) = -\infty$ for all $x \in \mathcal{X}$, so the supremum in (25) can be restricted to functions f bounded from below. If f is such a function, then for all $y \in \mathcal{Y}$, it holds

$$K_c f(x) \leqslant 2f(y) + G(2F(x,y)) \leqslant 2f(y) + \sup_{u\in\mathcal{X},v\in\mathcal{Y}} G(2F(u,v)).$$

So optimizing over y, one gets

$$K_c f(x) \leqslant 2\inf f + \sup_{u\in\mathcal{X},v\in\mathcal{Y}} G(2F(u,v))$$

and so

$$\int K_c f(x)\,\mu(dx) - \int f(y)\,\nu(dy) \leqslant \inf f + \sup_{u\in\mathcal{X},v\in\mathcal{Y}} G(2F(u,v))$$

Therefore, if $\inf f < \mathcal{I}_c(\mu,\nu) - \sup_{u\in\mathcal{X},v\in\mathcal{Y}} G(2F(u,v)) - 1$, then

$$\int K_c f(x)\,\mu(dx) - \int f(y)\,\nu(dy) < \mathcal{I}_c(\mu,\nu) - 1,$$

and so f can be dropped from the supremum in (25). We thus conclude that the supremum in (25) can be restricted to functions f bounded from below by a. In the case, where G is non-increasing, this lower bound can be improved. Indeed, if $f(y_0) < 0$ for some $y_0 \in \mathcal{Y}$, then for all $\lambda > 0$ it holds

$$K_c f(x) \leqslant \lambda f(y_0) + G(\lambda F(x,y_0)) \to -\infty$$

as $\lambda \to +\infty$. So the supremum in (25) can be restricted in this case to nonnegative functions.

Now let us show that the supremum in (25) can be further restricted to functions f such that $f \leqslant b$, where $b = [G'(0)]_- \sup_{x\in\mathcal{X},y\in\mathcal{Y}} F(x,y)$ with $[x]_- = \max(-x;0)$. Let $f \in \mathcal{L}^1(\nu)$; define $A = \{y\in\mathcal{Y} : f(y) \leqslant b\}$ and for all $m \in \mathcal{M}(\mathcal{Y})$ write $m_A(dy) = \mathbf{1}_A(y)\,m(dy)$ and $m_{A^c}(dy) = \mathbf{1}_{A^c}(y)\,m(dy)$. Since $u \mapsto G(u) + [G'(0)]_- u$ is non-decreasing, for all $x \in \mathcal{X}$, it holds

$$\begin{aligned}
\int \min(f,b)\,dm + c(x,m) &= \int f\,dm_A + bm(A^c) + G\left(\int F(x,y)\,m_A(dy) + \int F(x,y)\,m_{A^c}(dy)\right)\\
&\geqslant \int f\,dm_A + bm(A^c) + G\left(\int F(x,y)\,m_A(dy)\right) - [G'(0)]_- \int F(x,y)\,m_{A^c}(dy)\\
&\geqslant \int f\,dm_A + G\left(\int F(x,y)\,m_A(dy)\right) + \left(b - [G'(0)]_- \sup_{x\in\mathcal{X},y\in\mathcal{Y}} F(x,y)\right) m(A^c)\\
&\geqslant K_c f(x)
\end{aligned}$$

and so, letting $\hat{f} = \min(f,b)$, one gets $K_c\hat{f} \geqslant K_c f$. On the other hand, since $\hat{f} \leqslant f$, it also holds $K_c\hat{f} \leqslant K_c f$ and so $K_c f = K_c\hat{f}$. Since,

$$\int K_c f\,d\mu - \int f\,d\nu \leqslant \int K_c\hat{f}\,d\mu - \int \hat{f}\,d\nu$$

one concludes that the supremum in (25) can be restricted to functions bounded from above by b. In particular, when G is non-decreasing, then $[G'(0)]_- = 0$ and one can restrict to nonpositive functions.

Let us now show the dual attainment. Consider a sequence $(g_n)_{n\geqslant 0}$ of elements of $\mathcal{B} = \{f \in \mathcal{L}^1(\nu) : a \leqslant f \leqslant b\}$ such that $\int K_c g_n\,d\mu - \int g_n\,d\nu \to \mathcal{I}_c(\mu,\nu)$. According to the Dunford-Pettis theorem (or the Banach-Alaoglu-Bourbaki theorem), one can extract from $(g_n)_{n\geqslant 0}$ a subsequence (still denoted $(g_n)_{n\geqslant 0}$) converging to some $g_\infty \in \mathcal{B}$ for the weak topology $\sigma(L^1, L^\infty)$: for all $h \in L^\infty(\nu)$, $\int g_n h\,d\nu \to \int g_\infty h\,d\nu$. Moreover, according to Mazur's Lemma, there exists a sequence $(f_n)_{n\geqslant 0}$ of the form $f_n = \sum_{i=0}^{N_n} \lambda_i^{(n)} g_{n+i}$ with $N_n \geqslant 0$, $\lambda_0^{(n)},\ldots,\lambda_{N_n}^{(n)} \geqslant 0$ and $\sum_{i=0}^{N_n}\lambda_i^{(n)} = 1$ such that f_n converges strongly in $L^1(\nu)$ to g_∞, as $n \to \infty$. Extracting a subsequence if necessary, one can further assume that $(f_n)_{n\geqslant 0}$ converges ν almost everywhere to g_∞. Let $\Phi : \mathcal{B} \to \mathbb{R} : f \mapsto \int K_c f\,d\mu - \int f\,d\nu$. It is easily seen that

$$K_c((1-t)f + tg) \geqslant (1-t)K_c f + tK_c g, \qquad \forall t\in[0,1], \forall f,g \in \mathcal{B}$$

and so Φ is concave. Therefore

$$\Phi(f_n) \geqslant \sum_{i=0}^{N_n} \lambda_i^{(n)} \Phi(g_{n+i}) \geqslant \inf_{k\geqslant n} \Phi(g_k) \to \mathcal{I}_c(\mu,\nu),$$

as $n \to \infty$, and so $\Phi(f_n) \to \mathcal{I}_c(\mu,\nu)$ as $n \to \infty$. Since $K_c f_n \leqslant G(0)$ for all $n \geqslant 0$, one can apply Fatou's Lemma

$$\mathcal{I}_c(\mu,\nu) = \limsup_{n\to+\infty} \int K_c f_n \, d\mu - \lim_{n\to+\infty} \int f_n \, d\nu \leqslant \int \limsup_{n\to+\infty} K_c f_n \, d\mu - \int g_\infty \, d\nu.$$

For all $m \in \mathcal{M}(\mathcal{Y})$ and $x \in \mathcal{X}$, it holds

$$K_c f_n(x) \leqslant \int f_n \, dm + c(x,m)$$

and so, applying Fatou's Lemma again, one gets

$$\limsup_{n\to+\infty} K_c f_n(x) \leqslant \int \limsup_{n\to+\infty} f_n \, dm + c(x,m), \qquad x \in \mathcal{X}$$

and so, optimizing over m yields $\limsup_{n\to+\infty} K_c f_n \leqslant K_c(\bar f)$, with $\bar f = \limsup_{n\to+\infty} f_n \in \mathcal{B}$, and so

$$\mathcal{I}_c(\mu,\nu) \leqslant \int K_c \bar f \, d\mu - \int g_\infty \, d\nu.$$

Finally, since f_n converges ν almost everywhere to g_∞, it holds $g_\infty = \bar f$ ν a.e. and so

$$\mathcal{I}_c(\mu,\nu) \leqslant \int K_c \bar f \, d\mu - \int \bar f \, d\nu,$$

which shows that $\bar f$ is a dual optimizer. Finally, let us show dual attainment when $\mathcal{Y}$ is a finite set and $G'(0) \geqslant -\infty$. According to what precedes, the supremum in (25) can be restricted to functions $f \geqslant a$. On the other hand, since $\int K_c f \, d\mu \leqslant G(0)$ one can further restrict the supremum in (25) to functions f such that $\int f \, d\nu \leqslant G(0)+1-\mathcal{I}_c(\mu,\nu) := a'$. Since $\mathcal{Y}$ is finite, the set $\mathcal{C} = \{f \in \mathcal{L}^1(\nu) : a \leqslant f, \int f \, d\nu \leqslant a'\}$ is compact. Reasoning as above, one shows that any maximizing sequence of the dual problem admits a subsequence converging to a dual optimizer, which completes the proof. □

The following result relates primal and dual optimizers (provided they exist).

Proposition 4.1. *Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ be such that $\mathcal{I}_c(\mu,\nu) < +\infty$ and suppose that $\bar q$ is a kernel minimizer of $\mathcal{I}_c(\mu,\nu)$ and that $\bar f \in \mathcal{L}^1(\nu)$ is a dual optimizer:*

$$\mathcal{I}_c(\mu,\nu) = I_c^\mu[\bar q] = \int K_c \bar f \, d\mu - \int \bar f \, d\nu.$$

Then, the following relation holds true: for μ almost all $x \in \mathcal{X}$,

$$G'\left(\int F(x,z)\, \bar q^x(dz)\right) F(x,y) + \bar f(y) \geqslant 0, \qquad \forall y \in \mathcal{Y}. \tag{27}$$

In particular, if $G'(0) = -\infty$, then $\bar q^x(\mathcal{Y}) > 0$ for μ almost all $x \in \mathcal{X}$. Moreover, equality holds in (27) for $\bar q^x$ almost all $y \in \mathcal{Y}$.

Proof. Since

$$\mathcal{I}_c(\mu,\nu) = \int K_c \bar f \, d\mu - \int \bar f \, d\nu \leqslant \int\int \bar f(y)\, \bar q^x(dy)\mu(dx) + \int c(x,\bar q^x)\, \mu(dx) - \int \bar f \, d\nu = I_c^\mu[\bar q] = \mathcal{I}_c(\mu,\nu),$$

one concludes that

$$\int \bar f(y)\, \bar q^x(dy) + c(x,\bar q^x) = K_c \bar f(x),$$

for μ almost every $x \in \mathcal{X}$. Fix some $x \in \mathcal{X}$ for which the equality holds. By definition of K_c we thus get that for all $t \in (0,1)$ and $m \in \mathcal{M}(\mathcal{Y})$ such that $\int |f| \, dm < +\infty$ it holds

$$\int \bar f(y)\, \bar q^x(dy) + c(x,\bar q^x) \leqslant (1-t)\int \bar f(y)\, \bar q^x(dy) + t\int \bar f(y)\, m(dy) + c(x,(1-t)\bar q^x + tm).$$

Sending $t \to 0$ yields to

$$\int \bar{f}(y)\,(\bar{q}^x - m)(dy) \leqslant G'\left(\int F(x,z)\,\bar{q}^x(dz)\right)\int F(x,y)\,(m-\bar{q}^x)(dy)$$

(with the convention $-\infty \times 0 = 0$ if $G'(0) = -\infty$ and $m = \bar{q}^x$). In particular, if $G'(0) = -\infty$, then $\bar{q}^x \neq 0$. Rearranging terms yields to

$$\int \bar{f}(y) + G'\left(\int F(x,z)\,\bar{q}^x(dz)\right) F(x,y)\,\bar{q}^x(dy) \leqslant \int \bar{f}(y) + G'\left(\int F(x,z)\,\bar{q}^x(dz)\right) F(x,y)\,m(dy). \tag{28}$$

If the function $\psi_x : \mathcal{Y} \to \mathbb{R} : y \mapsto \bar{f}(y) + G'\left(\int F(x,z)\,\bar{q}^x(dz)\right) F(x,y)$ takes a negative value at some point $y_o \in \mathcal{Y}$ then choosing $m = \lambda\delta_{y_o}$ with $\lambda > 0$ arbitrary large yields a contradiction. Therefore, (27) holds true. Since the function ψ_x is nonnegative, one has $\inf_{m\in\mathcal{M}(\mathcal{Y})} \int \psi_x(y)\,m(dy) = 0$. Therefore taking the infimum over m in (28) yields to $\int \psi_x(y)\,\bar{q}^x(dy) = 0$ and so $\psi_x(y) = 0$ for $\bar{q}^x$ almost all $y \in \mathcal{Y}$. □

In a one dimensional framework, we now draw from (27) monotonicity properties of the supports of $\bar{q}^x$, $x \in \mathcal{X}$.

Theorem 4.2. *Let $\mathcal{X}$ and $\mathcal{Y}$ be two compact subsets of $\mathbb{R}$, $\mu \in \mathcal{P}(\mathcal{X}), \nu \in \mathcal{P}(\mathcal{Y})$ and suppose that $\bar{q} \in \mathcal{Q}(\mu,\nu)$ is a kernel solution of the transport problem (3) with cost function (24). For all $x \in \mathcal{X}$, denote by $S_x \subset \mathcal{Y}$ the support of $\bar{q}^x$. Suppose also that $F : \mathbb{R}\times\mathbb{R} \to (0,+\infty)$ is twice continuously differentiable and such that*

$$\frac{\partial^2 \ln F(x,y)}{\partial x \partial y} < 0, \qquad \forall x,y \in \mathbb{R}. \tag{29}$$

(1) *If G is increasing on $\mathbb{R}^+$, then there exists $A \subset \mathcal{X}$ with $\mu(A) = 1$ such that*

$$x_1 < x_2, x_1, x_2 \in A \Rightarrow \forall y_1 \in S_{x_1}, \forall y_2 \in S_{x_2}, y_1 \leqslant y_2. \tag{30}$$

(2) *If G is decreasing on $\mathbb{R}^+$ and $G'(0) > -\infty$, then there exists $A \subset \mathcal{X}$ with $\mu(A) = 1$ such that*

$$x_1 < x_2, x_1, x_2 \in A \Rightarrow \forall y_1 \in S_{x_1}, \forall y_2 \in S_{x_2}, y_1 \geqslant y_2. \tag{31}$$

The same conclusion holds if $G'(0) = -\infty$ provided $\mathcal{Y}$ is a finite set.

If μ has no atoms and full support, then there exists a unique right-continuous map $\bar{T} : \mathcal{X} \to \mathcal{Y}$ which is non-decreasing when G is increasing and non-increasing when G is decreasing such that any kernel solution $\tilde{q} \in \mathcal{Q}(\mu,\nu)$ of the transport problem (3) can be written as

$$\tilde{q}^x(dy) = \tilde{N}(x)\delta_{\bar{T}(x)},$$

for μ almost all $x \in \mathcal{X}$, with $\tilde{N} : \mathcal{X} \to \mathbb{R}^+$ a density with respect to μ. If G is assumed to be strictly convex, the density $\tilde{N}$ is unique also.

If F is such that

$$\frac{\partial^2 \ln F(x,y)}{\partial x \partial y} > 0, \qquad \forall x,y \in \mathbb{R}.$$

then all the conclusions are reversed: (30) holds when G is decreasing, (31) holds when G is increasing, and the monotonicity of $\bar{T}$ is the opposite of that of G.

Note that the existence of a kernel solution in Proposition 4.1 or Theorem 4.2 is granted at least in the following two cases : $G'(+\infty) = +\infty$ (according to Theorem 2.2) or $G'(+\infty) < +\infty$ and $\mathcal{X}$ is finite (according to Theorem 1.1).

Proof. We do the proof only in the case where G is increasing, the other case being similar. According to Theorem 4.1, there exists a nonpositive bounded function $\bar{f}$ achieving equality in the dual formula for $\mathcal{I}_c(\mu,\nu)$. According to Proposition 4.1, there is a set $A \subset \mathcal{X}$ with $\mu(A) = 1$ and such that for all $x \in A$ it holds

$$G'\left(\int F(x,z)\,\bar{q}^x(dz)\right) F(x,y) + \bar{f}(y) \geqslant 0, \qquad \forall y \in \mathcal{Y}.$$

Denote by $\tilde{S}_x$ the set of $y \in \mathcal{Y}$ for which the inequality above is an equality. According to Proposition 4.1, we know that $q^x(\tilde{S}_x) = 1$ for all $x \in A$. Condition (29) easily implies the following monotonicity property for F: if $a_1 < a_2$ and $b_1 < b_2$ then

$$F(a_1,b_1)F(a_2,b_2) < F(a_1,b_2)F(a_2,b_1). \tag{32}$$

Let $x_1, x_2 \in A$ such that $x_1 < x_2$ and suppose that there exist $y_1 \in \tilde{S}_{x_1}$ and $y_2 \in \tilde{S}_{x_2}$ such that $y_2 < y_1$. Then, denoting by $U(x) = \int F(x,z)\,\bar{q}^x(dz)$, $x \in \mathcal{X}$, it holds

$$\begin{aligned}
G'(U(x_1))F(x_1,y_1) &= -\bar{f}(y_1)\\
G'(U(x_2))F(x_2,y_2) &= -\bar{f}(y_2)\\
G'(U(x_1))F(x_1,y_2) &\geqslant -\bar{f}(y_2)\\
G'(U(x_2))F(x_2,y_1) &\geqslant -\bar{f}(y_1).
\end{aligned}$$

Multiplying the last two inequalities (note that all quantities are nonnegative) one gets

$$G'(U(x_1))G'(U(x_2))F(x_1,y_2)F(x_2,y_1) \geqslant \bar{f}(y_1)\bar{f}(y_2) = G'(U(x_1))G'(U(x_2))F(x_1,y_1)F(x_2,y_2).$$

Since G is increasing the term $G'(U(x_1))G'(U(x_2))$ is positive and can be simplified yielding to

$$F(x_1,y_2)F(x_2,y_1) \geqslant F(x_1,y_1)F(x_2,y_2),$$

which contradicts (32) with $a_1 = x_1$, $a_2 = x_2$, $b_1 = y_2$ and $b_2 = y_1$. Therefore, the family of sets $(\tilde{S}_x)_{x\in A}$ satisfies the following property:

$$x_1 < x_2 \Rightarrow y_1 \leqslant y_2, \forall y_1 \in \tilde{S}_{x_1}, \forall y_2 \in \tilde{S}_{x_2}.$$

Since $\bar{q}^x(\tilde{S}_x) = 1$, it is clear that $\tilde{S}_x$ is dense in S_x and so the same property is satisfied by the family $(S_x)_{x\in A}$, which proves (30).

Let us now assume that μ has no atoms. Let $\mathrm{is}(A)$ be the set of isolated points of A. Since this set is at most countable and μ has no atoms, it holds $\mu(\mathrm{is}(A)) = 0$. Thus, letting $A' = A\backslash \mathrm{is}(A)$, one gets $\mu(A') = 1$. Consider the maps $T_-, T_+ : A \to \mathbb{R}$ defined by $T_-(x) = \inf S_x$ and $T_+(x) = \sup S_x$ for all $x \in A'$. According to (30), the maps $T_\pm$ are non-decreasing. Let $D \subset A'$ be the set of points where T_- or T_+ is discontinuous. It is well known that D is at most countable. It is clear that whenever $T_-(x) < T_+(x)$ then $x \in D$. Defining $A'' = A'\backslash D$, and $T(x) = T_-(x) = T_+(x)$ for $x \in A''$, one thus gets that $S_x = \{T(x)\}$ for all $x \in A''$ and so there exists some nonnegative number $\bar{N}(x)$ such that $\bar{q}^x = \bar{N}(x)\delta_{T(x)}$ for all $x \in A''$. Finally, defining $\bar{T}(x) = \inf\{T(y) : y \in A'', y > x\}$, $x \in \mathcal{X}$, yields a non-decreasing and right-continuous extension of T to the whole space $\mathcal{X}$.

To prove the uniqueness part, we use a classical reasoning which goes back to [11]. Suppose that $\tilde{q} \in \mathcal{Q}(\mu,\nu)$ is another kernel solution of the transport problem. According to what precedes, there exists a density $\tilde{N}$ and a non-decreasing right-continuous map $\tilde{T}$ such that $\tilde{q}^x = \tilde{N}(x)\delta_{\tilde{T}(x)}$ for μ almost all $x \in \mathcal{X}$. By convexity, the nonnegative kernel $\frac{1}{2}(\bar{q} + \tilde{q})$ is also a solution. Therefore, there exists yet another non-decreasing right-continuous map U such that

$$\frac{1}{2}\left(\bar{N}(x)\delta_{\bar{T}(x)} + \tilde{N}(x)\delta_{\tilde{T}(x)}\right) = \frac{1}{2}(\bar{q}^x + \tilde{q}^x) = \left(\frac{1}{2}\bar{N}(x) + \frac{1}{2}\tilde{N}(x)\right)\delta_{U(x)},$$

for μ almost all $x \in \mathcal{X}$. Therefore, $\tilde{T}(x) = \bar{T}(x) = U(x)$ for μ almost all $x \in \mathcal{X}$. Since $\mathcal{X}$ is the support of μ and $\tilde{T}, \bar{T}$ are right-continuous, the equality holds true for all $x \in \mathcal{X}$. Furthermore, by optimality it holds

$$\int G\left((\frac{1}{2}\bar{N}(x) + \frac{1}{2}\tilde{N}(x))F(x,\bar{T}(x))\right)\,\mu(dx) = \frac{1}{2}\int G\left(\bar{N}(x)F(x,\bar{T}(x))\right)\,\mu(dx) + \frac{1}{2}\int G\left(\tilde{N}(x)F(x,\tilde{T}(x))\right)\,\mu(dx)$$

and so, assuming that G is strictly convex, one gets $\bar{N}(x) = \tilde{N}(x)$ for μ almost every $x \in \mathcal{X}$, which completes the proof. □

5. The particular case of conical cost functions

This section is devoted to the study of the transport problem (3), when c is a conical cost function. We will first obtain an improved duality result showing in particular that under mild conditions there is dual attainment. Then we will obtain a Strassen type result for a variant of the convex order involving positively 1-homogenous functions. Finally, we will prove structure results for primal and dual solutions.

5.1. **Framework.** In the whole section, we adopt the following framework :

- $\mathcal{X}$ is a compact metrizable space,
- $\mathcal{Y}$ is a compact subset of $\mathbb{R}^d$, equipped with some arbitrary norm $\|\cdot\|$, and we denote by $\mathrm{co}(\mathcal{Y})$ its convex hull and by $\mathcal{Z}$ its conical hull, i.e

$$\mathcal{Z} = \left\{ \sum_{i=1}^{n} \lambda_i y_i : \lambda_1, \ldots, \lambda_n \in \mathbb{R}_+, y_1, \ldots, y_n \in \mathcal{Y}, n \geqslant 1 \right\},$$

- the cost function $c : \mathcal{X} \times \mathcal{M}(\mathcal{Y}) \to \mathbb{R}$ is of the following form

$$c(x,m) = F\left(x, \int y\,dm\right), \qquad x \in \mathcal{X}, m \in \mathcal{M}(\mathcal{Y}), \tag{33}$$

where $F : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$ is lower semicontinuous on $\mathcal{X} \times \mathcal{Z}$ and convex with respect to its second variable.

When c is of this form we will say that c is a *conical cost function*.

First let us translate Assumptions (A), (B) and (C) in this framework. Let us introduce the recession function of F, defined by

$$F'_\infty(x,z) = \lim_{\lambda \to +\infty} \frac{F(x,\lambda z)}{\lambda}, \qquad x \in \mathcal{X}, z \in \mathcal{Z}.$$

- Assumption (A) is fulfilled as soon as F satisfies the following condition (A'): there exists a family of continuous functions $(a_k)_{k \geqslant 0}$ on $\mathcal{X}$ and a family of continuous functions $(u_k)_{k \geqslant 0}$ on $\mathcal{X}$ with values in $\mathbb{R}^d$ such that

$$F(x,z) = \sup_{k \geqslant 0} \{u_k(x) \cdot z + a_k(x)\}, \qquad x \in \mathcal{X}, z \in \mathcal{Z}. \tag{A'}$$

Note that if F satisfies (A'), then the corresponding cost function c satisfies (A) with $b_k(x,y) = u_k(x) \cdot y$ and the same a_k for $k \geqslant 0$.
- Assumption (B) is satisfied by c as soon as F satisfies the following condition (B')

$$F'_\infty(x,z) = +\infty, \qquad \forall x \in \mathcal{X}, \forall z \in \mathcal{Z}\backslash\{0\}. \tag{B'}$$

- Finally, Assumption (C) holds for c as soon as F satisfies the following condition (C') : for all $z \in \mathcal{Z}$, the functions $x \mapsto F(x,z)$ and $x \mapsto F'_\infty(x,z)$ are continuous on $\mathcal{X}$ and there exists $a \in \mathbb{R}_+$ such that $F'_\infty(x,z) \leqslant a$ for all $x \in \mathcal{X}$ and $z \in \mathrm{co}(\mathcal{Y})$.

5.2. **Duality and dual attainment for conical cost functions.** The following result improves the conclusion of Theorem 3.2 in the case of conical cost functions. Recall that a function $\varphi : \mathcal{Z} \to \mathbb{R} \cup \{+\infty\}$ defined on a cone $\mathcal{Z} \subset \mathbb{R}^d$ is positively 1-homogenous if $\varphi(\lambda x) = \lambda\varphi(x)$, for all $\lambda \geqslant 0$ and for all $x \in \mathcal{Z}$ such that $\varphi(x) < +\infty$.

Theorem 5.1. *With the above notation, and further assuming that*

- 0 *does not belong to* $\mathrm{co}(\mathcal{Y})$,
- *there exists* $\lambda > 1$ *such that* $M := \sup_{y\in\mathcal{Y}} \int F(x, \lambda y)\, \mu(dx) < +\infty$,
- F *satisfies* (A'),

then it holds

$$\mathcal{I}_c(\mu, \nu) = \sup_{\varphi\in\Phi(\mathcal{Z})\cap L^1(\nu)} \left\{ \int Q_F\varphi(x)\, \mu(dx) - \int \varphi(y)\, \nu(dy) \right\}, \qquad \forall \nu \in \mathcal{P}(\mathcal{Y}) \tag{34}$$

where $\Phi(\mathcal{Z})$ *is the set of all lower semicontinuous, convex positively* 1*-homogenous functions* $\varphi : \mathcal{Z} \to \mathbb{R} \cup \{+\infty\}$ *and where*

$$Q_F\varphi(x) = \inf_{z\in\mathcal{Z}} \{\varphi(z) + F(x, z)\}, \qquad x \in \mathcal{X}.$$

Moreover, one can further restrict the supremum in (34) *to functions* $\varphi \in \Phi(\mathcal{Z}) \cap L^1(\nu)$ *such that*

$$\varphi(z) \geqslant -\frac{M}{\lambda - 1}, \qquad \forall z \in \mathrm{co}(\mathcal{Y}). \tag{35}$$

Furthermore, there exists a function $\bar{\varphi} \in \Phi(\mathcal{Z}) \cap L^1(\nu)$ *satisfying* (35) *and such that*

$$\mathcal{I}_c(\mu, \nu) = \int Q_F\bar{\varphi}(x)\, \mu(dx) - \int \bar{\varphi}(y)\, \nu(dy).$$

Proof. The proof is divided into three steps.
Step 1. In this step, we show that

$$\sup_{\varphi\in\Phi(\mathcal{Z})\cap L^1(\nu)} \int Q_F\varphi\, d\mu - \int \varphi\, d\nu \leqslant \mathcal{I}_c(\mu, \nu). \tag{36}$$

If $\varphi \in \Phi(\mathcal{Z})$ is ν integrable and $q \in \mathcal{Q}(\mu, \nu)$, then using Jensen's inequality and the positive 1-homogeneity of φ, it holds

$$\int \varphi\, d\nu = \int \left(\int \varphi(y)\, q^x(dy) \right) \mu(dx) \geqslant \int \varphi \left(\int y\, q^x(dy) \right) \mu(dx).$$

On the other hand

$$\begin{aligned} \int Q_F\varphi\, d\mu &\leqslant \int \varphi\left(\int y\, q^x(dy) \right) + F\left(x, \int y\, q^x(dy)\right) \mu(dx) \\ &\leqslant \int \varphi\, d\nu + \int F\left(x, \int y\, q^x(dy)\right) \mu(dx). \end{aligned}$$

Thus optimizing over φ and over q gives (36).

Step 2. In this step, we assume that c satisfies Assumption (B') and we prove that the converse inequality holds true in (36) and that the supremum can be restricted to functions satisfying (35). Recall that according to Theorem 3.2, it holds

$$\mathcal{I}_c(\mu, \nu) = \sup_{f\in\mathcal{C}_b(\mathcal{Y})} \int K_c f(x)\, \mu(dx) - \int f(y)\, \nu(dy).$$

Observe that for all $f \in \mathcal{C}_b(\mathcal{Y})$

$$K_c f(x) = \inf_{m \in \mathcal{M}(\mathcal{Y})} \left\{ \int f(y)\, m(dy) + F\left(x, \int y\, m(dy)\right) \right\} = Q_F \bar{f}(x),$$

where

$$\bar{f}(z) = \inf \left\{ \int f(y)\, m(dy) : m \in \mathcal{M}(\mathcal{Y}), \int y\, m(dy) = z \right\}, \qquad z \in \mathcal{Z}.$$

Note that $\bar{f}$ is the greatest convex and positively 1-homogenous function $\varphi : \mathcal{Z} \to \mathbb{R}$ such that $\varphi \leqslant f$ on $\mathcal{Y}$. Furthermore, the function $\bar{f}$ is lower semicontinuous and so $\bar{f} \in \Phi(\mathcal{Z})$. Indeed, if $z_n \to z$ in $\mathcal{Z}$ then there exists a sequence $m_n \in \mathcal{M}(\mathcal{Y})$ such that $\int y\, m_n(dy) = z_n$ and $\bar{f}(z_n) = \int f(y)\, m_n(dy) + \epsilon_n$, with $\epsilon_n \to 0$. Let $D = \sup_{n \geqslant 0} \left\| \int y\, m_n(dy) \right\| < +\infty$. Since 0 does not belong to the compact convex set $\mathrm{co}(\mathcal{Y})$, there is $\alpha > 0$ such that $\|y\| \geqslant \alpha$ for all $y \in \mathrm{co}(\mathcal{Y})$. Thus, one gets

$$\alpha m_n(\mathcal{Y}) \leqslant \left\| \int y\, m_n(dy) \right\| \leqslant D.$$

One concludes from Prokhorov's theorem that the sequence m_n admits a converging subsequence. The claim then follows from the lower semicontinuity of the function $m \mapsto \int f(z)\, m(dz)$.

By construction $\bar{f} \leqslant f$ on $\mathcal{Y}$ and since $\bar{f}$ is convex, it is bounded from below by some affine map. Thus $\bar{f}$ is ν-integrable. Also, since $\bar{f} \leqslant f$, it holds

$$\int K_c f(x)\, \mu(dx) - \int f(y)\, \nu(dy) \leqslant \int Q_F \bar{f}(x)\, \mu(dx) - \int \bar{f}(y)\, \nu(dy).$$

Optimizing, one thus gets that

$$\mathcal{I}_c(\mu, \nu) \leqslant \sup_{\varphi \in \Phi(\mathcal{Z}) \cap L^1(\nu)} \int Q_F \varphi(x)\, \mu(dx) - \int \varphi(y)\, \nu(dy),$$

which completes the proof of (34).

Now, let us show that the supremum can be restricted to functions $\varphi \in \Phi(\mathcal{Z}) \cap L^1(\nu)$ satisfying (35). If $\varphi \in \Phi(\mathcal{Z}) \cap L^1(\nu)$, then being ν integrable, it takes at least one finite value on the support of ν. Since φ is also lower semicontinuous, it reaches its infimum on $\mathrm{co}(\mathcal{Y})$ at some point $z_0 \in \mathrm{co}(\mathcal{Y})$. By definition of $Q_F \varphi$, it holds

$$Q_F \varphi(x) \leqslant \varphi(\lambda z_0) + F(x, \lambda z_0), \qquad \forall x \in \mathcal{X}.$$

Therefore, φ being positively homogenous, it holds

$$\int Q_F \varphi(x)\, \mu(dx) - \int \varphi(y)\, \nu(dy) \leqslant (\lambda - 1)\varphi(z_0) + \int F(x, \lambda z_0)\, \mu(dx) \leqslant (\lambda - 1)\varphi(z_0) + M.$$

Thus, if $\varphi(z_0) < -\frac{M}{\lambda - 1}$, then $\int Q_F \varphi(x)\, \mu(dx) - \int \varphi(y)\, \nu(dy) < 0$. Since $\mathcal{I}_c(\mu, \nu) \geqslant 0$, one can thus drop such functions φ from the supremum in (34), which completes the proof.

Step 3. In this step, we remove the assumption that c satisfies (B') and we prove dual attainment. Without loss of generality, we will assume that $\mathcal{Y} = \mathrm{co}\,\mathrm{Supp}(\nu)$. Note that this is always possible since the convex hull of a compact set is itself compact. To make the proof easier to read, we will also assume that $\mathcal{Y} = \mathrm{co}\,\mathrm{Supp}(\nu)$ has a non empty interior. If this is not the case, then one can easily adapt the arguments below using the notions of relative interior and relative boundary of a convex set.

For all $n \geqslant 1$, define

$$F_n(x, z) = F(x, z) + \frac{1}{n}\|z\|^2, \qquad x \in \mathcal{X}, z \in \mathcal{Z},$$

and $c_n(x, m) = F_n(x, \int y\, dm)$, $x \in \mathcal{X}$, $m \in \mathcal{M}(\mathcal{Y})$. It is clear that F_n satisfies both (A') and (B').

Observe that for all $n \geqslant 1$, $\mathcal{I}_c(\mu, \nu) \leqslant \mathcal{I}_{c_n}(\mu, \nu)$.

Let $(\varphi_n)_{n\geqslant 1}$ be a sequence in $\Phi(\mathcal{Z})\cap L^1(\nu)$ satisfying (35) and such that for all $n\geqslant 1$

$$\mathcal{I}_{c_n}(\mu,\nu)\leqslant \int Q_{F_n}\varphi_n\,d\mu-\int \varphi_n\,d\nu+\frac{1}{n}.$$

Such a sequence exists thanks to Step 2. For all $n\geqslant 1$, it holds

$$Q_{F_n}\varphi_n(x)\leqslant \varphi_n(0)+F_n(x,0)=F(x,0),\qquad \forall x\in\mathcal{X}. \tag{37}$$

Therefore, using the integrability condition on μ, one gets that $\sup_{n\geqslant 1}\int Q_{F_n}\varphi_n(x)\,\mu(dx)<+\infty$. Since $\int Q_{F_n}\varphi_n\,d\mu-\int\varphi_n\,d\nu\geqslant \mathcal{I}_c(\mu,\nu)-\frac{1}{n}$, this implies in particular that $\sup_{n\geqslant 1}\int\varphi_n\,d\nu<+\infty$.

Let us show that $(\varphi_n)_{n\geqslant 1}$ admits a converging subsequence. Define $\widetilde{\varphi}_n=\varphi_n+\frac{M}{\lambda-1}$, $n\geqslant 1$. For all $n\geqslant 1$, $p\geqslant 0$, $y_1,\ldots,y_p\in\operatorname{Supp}(\nu)$ and $\lambda_1,\ldots,\lambda_p\geqslant 0$ such that $\sum_{i=1}^p\lambda_i=1$, it follows from Jensen's inequality and the fact that $\widetilde{\varphi}_n\geqslant 0$ on $\mathcal{Y}$ that

$$\begin{aligned}\widetilde{\varphi}_n\left(\sum_{i=1}^p\lambda_i\frac{\int_{B(y_i,\epsilon)}z\,\nu(dz)}{\nu(B(y_i,\epsilon))}\right)\leqslant \sum_{i=1}^p\lambda_i\widetilde{\varphi}_n\left(\frac{\int_{B(y_i,\epsilon)}z\,\nu(dz)}{\nu(B(y_i,\epsilon))}\right)&\leqslant \sum_{i=1}^p\lambda_i\frac{1}{\nu(B(y_i,\epsilon))}\int_{B(y_i,\epsilon)}\widetilde{\varphi}_n(z)\,\nu(dz)\\ &\leqslant\left(\sum_{i=1}^p\lambda_i\frac{1}{\nu(B(y_i,\epsilon))}\right)\int\widetilde{\varphi}_n(z)\,\nu(dz),\end{aligned}$$

where $B(y,\epsilon)$ denotes the open ball centered at y of radius $\epsilon>0$. Since $\sup_{n\geqslant 1}\int\widetilde{\varphi}_n\,d\nu<+\infty$, it holds

$$\sup_{n\geqslant 1}\widetilde{\varphi}_n(u)<+\infty$$

for all u belonging to the set $C=\operatorname{co}\left\{\frac{\int_{B(y,\epsilon)}z\,\nu(dz)}{\nu(B(y,\epsilon))}:y\in\operatorname{Supp}(\nu),\epsilon>0\right\}$. Let us show that C is dense in $\mathcal{Y}$. Take $y\in\mathcal{Y}$; since $\mathcal{Y}=\operatorname{co}\operatorname{Supp}(\nu)$, there exists $y_1,\ldots,y_p\in\operatorname{Supp}(\nu)$ and $\lambda_1,\ldots,\lambda_p\geqslant 0$ such that $\sum_{i=1}^p\lambda_i=1$ and $y=\sum_{i=1}^p\lambda_iy_i$. For all $\epsilon>0$, define $y_\epsilon=\sum_{i=1}^p\lambda_i\frac{\int_{B(y_i,\epsilon)}z\,\nu(dz)}{\nu(B(y_i,\epsilon))}\in C$. Then it is easily seen that $y_\epsilon\to y$ when $\epsilon\to 0$, which proves the claim.

According to [32, Theorem 10.9 page 91], one can extract from $(\widetilde{\varphi}_n)_{n\geqslant 1}$ a subsequence (we will still denote it by $(\widetilde{\varphi}_n)_{n\geqslant 1}$ not to overload the notation) converging pointwise on $\operatorname{int}(\mathcal{Y})$ to some convex function. Of course, the sequence $(\varphi_n)_{n\geqslant 1}$ also converges pointwise on $\operatorname{int}(\mathcal{Y})$. Since $\mathcal{Z}=\mathbb{R}_+\mathcal{Y}$ and $\operatorname{int}(\mathcal{Z})=\mathbb{R}_+^*\operatorname{int}(\mathcal{Y})$, the positive homogeneity of the functions φ_n implies that φ_n converges pointwise on $\operatorname{int}(\mathcal{Z})$. Set $\varphi(x)=\lim_{n\to\infty}\varphi_n(x)$, for all $x\in\operatorname{int}(\mathcal{Z})$. Extend φ by setting $\varphi(a)=\liminf_{z\to a,z\in\operatorname{int}(\mathcal{Z})}\varphi(z)$ whenever $a\in\partial\mathcal{Z}$, so that φ is lower semicontinuous on $\mathcal{Z}$ (and still convex and positively homogenous). If $a\in\partial\mathcal{Z}$ and $z\in\operatorname{int}(\mathcal{Z})$, then for all $t\in(0,1)$ the point $(1-t)a+tz$ belongs to $\operatorname{int}(\mathcal{Z})$. Therefore, letting $n\to\infty$ in the inequality

$$\varphi_n((1-t)a+tz)\leqslant(1-t)\varphi_n(a)+t\varphi_n(z)$$

one gets that

$$\varphi((1-t)a+tz)\leqslant(1-t)\liminf_{n\to+\infty}\varphi_n(a)+t\varphi(z)$$

and letting $t\to 0$ gives then

$$\varphi(a)\leqslant\liminf_{n\to+\infty}\varphi_n(a).$$

Then,

$$\begin{aligned}\mathcal{I}_c(\mu,\nu) &\leqslant \limsup_{n\to+\infty}\left(\int Q_{F_n}\varphi_n\,d\mu - \int \varphi_n\,d\nu\right)\\ &\leqslant \limsup_{n\to+\infty}\int Q_{F_n}\varphi_n\,d\mu - \liminf_{n\to+\infty}\int \varphi_n\,d\nu\\ &\leqslant \int \limsup_{n\to+\infty} Q_{F_n}\varphi_n\,d\mu - \int \liminf_{n\to+\infty}\varphi_n\,d\nu\\ &\leqslant \int \limsup_{n\to+\infty} Q_{F_n}\varphi_n\,d\mu - \int \varphi\,d\nu,\end{aligned}$$

where the third inequality follows by Fatou lemma (note that, for all $n \geqslant 1$, $\varphi_n \geqslant -M/(\lambda-1)$ on $\mathcal{Y}$ and $Q_{F_n}\varphi_n \leqslant F(\,\cdot\,,0)$ which is μ-integrable), and the last inequality from the fact that $\liminf_{n\to+\infty}\varphi_n \geqslant \varphi$ on $\mathcal{Z}$. For all $n \geqslant 0$, it holds

$$\varphi_n(z) \geqslant Q_F\varphi_n(x) - F_n(x,z), \qquad \forall x \in \mathcal{X}, \forall z \in \mathcal{Z}.$$

Therefore, letting $n \to \infty$, one gets that

$$\varphi(z) \geqslant \limsup_{n\to+\infty} Q_{F_n}\varphi_n(x) - F(x,z), \qquad \forall x \in \mathcal{X}, \forall z \in \operatorname{int}(\mathcal{Z}).$$

This inequality is still true when $z \in \partial\mathcal{Z}$. Indeed, fix $z \in \partial\mathcal{Z}$ and $z' \in \operatorname{int}(\mathcal{Z})$; since φ and $F(x,\cdot\,)$ are both convex and lower semicontinuous on $\mathcal{Z}$, they satisfy $\varphi(z) = \lim_{t\to 0}\varphi((1-t)z+tz')$ and, for all $x \in \mathcal{X}$, $F(x,z) = \lim_{t\to 0} F(x,(1-t)z+tz')$. Since $(1-t)z+tz' \in \operatorname{int}(\mathcal{Z})$, this easily implies that the inequality above is also true for z. From this follows that

$$\limsup_{n\to+\infty} Q_{F_n}\varphi_n(x) \leqslant \inf_{z\in\mathcal{Z}}\{\varphi(z)+F(x,z)\} = Q_F\varphi(x), \qquad \forall x \in \mathcal{X}.$$

In conclusion, we have shown the existence of a function $\varphi \in \Phi(\mathcal{Z}) \cap L^1(\nu)$ such that

$$\mathcal{I}_c(\mu,\nu) \leqslant \int Q_F\varphi\,d\mu - \int \varphi\,d\nu.$$

Since the converse inequality is always true (according to Step 1), this completes the proof. □

5.3. **A new variant of Strassen Theorem.** Recall that if μ,ν are two probability measures on $\mathbb{R}^d$ having a finite first moment, μ is said to be dominated by ν in the convex order, if

$$\int \varphi\,d\mu \leqslant \int \varphi\,d\nu,$$

for all convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$. In this case, we denote this relation by $\mu \leqslant_c \nu$. According to a well known result due to Strassen [33], $\mu \leqslant_c \nu$ if and only if there exists a martingale coupling with marginals μ and ν, that is to say a couple (U,V) of random vectors with $U \sim \mu$, $V \sim \nu$ and $\mathbb{E}[V \mid U] = U$ a.s.

Transport problems with conical cost functions introduced above are naturally related to the following variant of the convex order. If μ,ν are two probability measures with a finite moment of order 1, we will say that μ is dominated by ν for the positively 1-homogenous convex order if for all $\varphi : \mathbb{R}^d \to \mathbb{R}$ convex and positively 1-homogenous, it holds $\int \varphi\,d\mu \leqslant \int \varphi\,d\nu$. We will use the notation $\mu \leqslant_{phc} \nu$ to denote this order.

The following result generalizes Strassen's theorem to this restricted convex order. Note that if ν is compactly supported, then $\mu \leqslant_{phc} \nu$ does not imply that μ is also compactly supported.

Theorem 5.2. *Let μ,ν be two probability measures on $\mathbb{R}^d$ and suppose that μ has a finite moment of order 1 and that ν is compactly supported and such that the convex hull of its support does not contain 0. The following are equivalent* (*i*) $\mu \leqslant_{phc} \nu$,

(*ii*) *There exists a nonnegative kernel* q *such that* $\mu q = \nu$ *and*

$$\int y\, q^x(dy) = x$$

for μ *almost every* x,

(*iii*) *There exists a probability measure* η *absolutely continuous with respect to* μ *with density denoted by* N *and a couple of random vectors* (U, V) *with* $U \sim \eta$, $V \sim \nu$ *such that*

$$N(U)\mathbb{E}[V \mid U] = U \text{ a.s.} \tag{38}$$

Note that (38) also means that $(U, N(U)V)$ is a martingale.

Remark 5.1. *Note that in dimension* 1, *the conclusion of Theorem 5.2 is essentially trivial. Indeed, it is easy to see that* $\mu \leqslant_{phc} \nu$ *if and only if* $\int x\, d\mu = \int x\, d\nu$, $\int [x]_+\, d\mu \leqslant \int [x]_+\, d\nu$ *and* $\int [x]_-\, d\mu \leqslant \int [x]_-\, d\nu$. *By assumption, the convex hull of the support of* ν *does not contain* 0, *so it is contained either in* $(0, \infty)$ *or in* $(-\infty, 0)$. *Let us assume without loss of generality that the support of* ν *is contained in* $(0, \infty)$. *Then it holds* $\int [x]_-\, d\mu \leqslant \int [x]_-\, d\nu = 0$ *and so the support of* μ *is also contained in* $(0, \infty)$. *Consider the nonnegative function* $N(x) = \frac{x}{\int x\, d\mu} \mathbf{1}_{(0,\infty)}(x)$, *which satisfies* $\int N(x)\, \mu(dx) = 1$, *and define* $\eta = N\, \mu$. *Let* U, V *be two independent random variables such that* $U \sim \eta$ *and* $V \sim \nu$, *then it holds*

$$\mathbb{E}[V \mid U] = \mathbb{E}[V] = \int x\, d\mu = \frac{U}{N(U)} \text{ a.s.}$$

In higher dimension, it is not clear if such simple and explicit construction is available.

Theorem 5.2 will follow from the following slightly more general result where the assumption that 0 does not belong to the convex hull of the support of ν is removed (but μ is compactly supported).

Theorem 5.3. *Let* μ, ν *be two compactly supported probability measures on* $\mathbb{R}^d$. *The following are equivalent:*

(*i*) $\mu \leqslant_{phc} \nu$,

(*ii*) *The probability measure* ν *can be decomposed as the sum of two nonnegative measures* $\nu = \nu_1 + \nu_2$ *such that* $\int y\, \nu_2(dy) = 0$ *and there exists a nonnegative kernel* $q \in \mathcal{Q}(\mu, \nu_1)$ *such that*

$$\int y\, q^x(dy) = x$$

for μ *almost every* x.

Proof of Theorem 5.3. Let us show that (ii) implies (i). Let φ be some convex positively 1-homogenous function ; according to Jensen's inequality and positive 1-homogeneity, it holds

$$\int \varphi(x)\, \mu(dx) = \int \varphi\left(\int y\, q^x(dy)\right) \mu(dx) \leqslant \iint \varphi(y)\, q^x(dy)\, \mu(dx) = \int \varphi(y)\, \nu_1(dy) \leqslant \int \varphi(y)\, \nu(dy),$$

where the last inequality comes from the fact that $\int \varphi(y)\, \nu_2(dy) \geqslant \varphi\left(\int y\, \nu_2(dy)\right) = 0$. Now let us show that (i) implies (ii). Let us denote by $\mathcal{X}$ and $\mathcal{Y}$ the compact supports of μ and ν and consider the cost function $c : \mathcal{X} \times \mathcal{M}(\mathcal{Y}) \to \mathbb{R}_+$

$$c(x, m) = F\left(x, \int y\, m(dy)\right) = \left\| x - \int y\, m(dy) \right\|, \qquad \forall x \in \mathcal{X}, \forall m \in \mathcal{M}(\mathcal{Y}),$$

with $F(x, z) = \|x - z\|$, $x, z \in \mathbb{R}^d$. It is not difficult to see that F satisfies Assumption (A'). Therefore, according to Theorem 5.1, the following duality formula holds

$$\mathcal{I}_c(\mu, \nu) = \sup_{\varphi \in \Phi(\mathcal{Z}) \cap L^1(\nu)} \left\{ \int Q\varphi(x)\, \mu(dx) - \int \varphi(y)\, \nu(dy) \right\},$$

with $Q\varphi(x) = \inf_{z\in\mathcal{Z}}\{\varphi(z) + \|x-z\|\}$, $x \in \mathbb{R}^d$. The supremum can be restricted to φ that are bounded from below by some constant $\kappa \in \mathbb{R}$. For such functions φ, it is easy to check that $Q\varphi$ is finite valued, convex and positively 1-homogenous on $\mathbb{R}^d$. Thus $\int Q\varphi\, d\mu \leqslant \int Q\varphi\, d\nu$. Since $Q\varphi \leqslant \varphi$ on $\mathcal{Z}$, we conclude that $\int Q\varphi\, d\mu \leqslant \int \varphi\, d\nu$. Therefore, $\mathcal{I}_c(\mu,\nu) = 0$. On the other hand, since F satisfies Assumption (C'), it follows from Theorem 2.1 and Lemma 2.4 that there exists some $\pi \in \Pi(\mathrm{Supp}(\mu),\nu)$ such that

$$\bar{I}_c^\mu[\pi] = \bar{\mathcal{I}}_c(\mu,\nu) = \mathcal{I}_c(\mu,\nu) = 0.$$

Since $c'_\infty(x,m) = \|\int y\, m(dy)\|$, $m \in \mathcal{M}(\mathcal{Y})$, one thus gets

$$\int \left\| x - \frac{d\eta^{ac}}{d\mu}(x)\int y\,\pi^x(dy)\right\| \mu(dx) + \int \left\| \int y\,\pi^x(dy)\right\| \eta^s(dx) = 0, \tag{39}$$

where $\eta = \eta^{ac} + \eta^s$ denotes the first marginal of π. Let us define $q^x(dy) = \frac{d\eta^{ac}}{d\mu}(x)\pi^x(dy)$, $x \in \mathcal{X}$, $\nu_1 = \mu q$ and $\nu_2 = \nu - \nu_1$. It follows from (39) that $\int y\, q^x(dy) = x$ for μ almost all x. Moreover, $\int y\,\pi^x(dy) = 0$ for η^s almost all x. Therefore,

$$\int y\,\nu_2(dy) = \int\left(\int y\pi^x(dy)\right)\eta(dx) - \int\left(\int y\pi^x(dy)\right)\eta^{ac}(dx) = \int\left(\int y\pi^x(dy)\right)\eta^s(dx) = 0,$$

which completes the proof of (ii). □

Proof of Theorem 5.2. It is clear that (ii) and (iii) are equivalent. The proof of $(ii) \Rightarrow (i)$ is exactly the same as the one of $(ii) \Rightarrow (i)$ in Theorem 5.3.

Let us now prove that $(i) \Rightarrow (ii)$. First assume that μ is compactly supported. According to Theorem 5.3, the probability measure ν can be written as $\nu = \nu_1 + \nu_2$ with $\int y\,\nu_2(dy) = 0$ and there exists $q \in \mathcal{Q}(\mu,\nu_1)$ satisfying $\int y\, q^x(dy) = x$ for μ almost all x. If $\nu_2(\mathcal{Y}) \neq 0$, then $0 = \frac{\int_{\mathcal{Y}} y\,\nu_2(dy)}{\nu_2(\mathcal{Y})} \in \mathrm{co}(\mathcal{Y})$, which contradicts our assumptions. Thus $\nu_2(\mathcal{Y}) = 0$ and so $\nu_1 = \nu$, which proves (ii), when μ has a compact support.

Now let us relax the assumption that μ has a compact support. Let us construct a sequence of compactly supported probability measures $(\mu_n)_{n\geqslant 1}$ converging to μ in the weak topology and such that $\mu_n \leqslant_c \mu$ for all $n \geqslant 1$. One can for instance obtain such a sequence as follows. Consider $C_n = [-n,n]^d$ and write $\mathbb{R}^d\backslash C_n = \bigcup_{1\leqslant k\leqslant K_n} D_{n,k}$, where $(D_{n,k})_{1\leqslant k\leqslant K_n}$ are disjoints convex subsets of $\mathbb{R}^d$. Then define

$$\mu_n(dx) = \mathbf{1}_{C_n}(x)\,\mu(dx) + \sum_{k=1}^{K_n}\mu(D_{n,k})\delta_{x_{n,k}}(dx)$$

where $x_{n,k} = \frac{1}{\mu(D_{n,k})}\int_{D_{n,k}} x\,\mu(dx)$ if $\mu(D_{n,k}) > 0$ and any point in $D_{n,k}$ otherwise. If $f : \mathbb{R}^d \to \mathbb{R}$ is a convex function, then it follows from Jensen's inequality that

$$\int f(x)\,\mu_n(dx) = \int_C f(x)\,\mu(dx) + \sum_{k=1}^{K_n}\mu(D_{n,k})f(x_{n,k}) \leqslant \int_C f(x)\,\mu(dx) + \sum_{k=1}^{K_n}\int_{D_{n,k}} f(x)\,\mu(dx) = \int f(x)\,\mu(dx)$$

and so $\mu_n \leqslant_c \mu$. Also, for any bounded continuous function $f : \mathbb{R}^d \to \mathbb{R}$, it holds

$$\left|\int f\,d\mu_n - \int f\,d\mu\right| = \left|\int_{\mathbb{R}^d\backslash C_n} f\,d\mu_n - \int_{\mathbb{R}^d\backslash C_n} f\,d\mu\right| \leqslant 2\|f\|_\infty\,(1-\mu(C_n)) \to 0,$$

as $n \to \infty$, and so $(\mu_n)_{n\geqslant 1}$ converges to μ in the weak topology.

Since $\mu_n \leqslant_c \mu$ and $\mu \leqslant_{phc} \nu$, it is clear that $\mu_n \leqslant_{phc} \nu$. By construction, μ_n has a compact support, and so there exists a nonnegative kernel q_n such that $\mu_n q_n = \nu$ and $\int y\, q_n^x(dy) = x$ for μ_n almost all $x \in \mathbb{R}^d$. For all $n \geqslant 1$, write for all $x \in \mathbb{R}^d$, $q_n^x = N_n(x)p_n^x$ where p_n^x is a probability measure and denote $\eta_n(dx) = N_n(x)\mu_n(dx)$ and $\pi_n(dxdy) = \eta_n(dx)p_n^x(dy)$. Let us show that the sequence $(\eta_n)_{n\geqslant 1}$

is tight. By assumption, $\int \|x\| \, \mu(dx) < +\infty$; thus by the de la Vallée Poussin theorem, there exists some non-decreasing convex function $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ such that $\alpha(0) = 0$, $\alpha(x)/x \to +\infty$ as $x \to +\infty$ and $\int \alpha(\|x\|) \, \mu(dx) < +\infty$. The function $x \mapsto \alpha(\|x\|)$ being convex, we thus get that

$$\sup_{n \geqslant 1} \int \alpha(\|x\|) \, \mu_n(dx) \leqslant \int \alpha(\|x\|) \, \mu(dx) := M$$

and, since $\int y \, q_n^x(dy) = x$ for μ_n almost all x,

$$\sup_{n \geqslant 1} \int \alpha \left(\left\| \int y \, q_n^x(dy) \right\| \right) \mu_n(dx) \leqslant M$$

But,

$$\int \alpha \left(\left\| \int y \, q_n^x(dy) \right\| \right) \mu_n(dx) = \int \alpha \left(N_n(x) B_n(x) \right) \mu_n(dx),$$

where $B_n(x) = \left\| \int y \, p_n^x(dy) \right\|$. Since $\int y \, p_n^x(dy)$ belongs to the convex hull of the support of ν which is a compact convex set not containing zero, there exists some $b > 0$ independent of n such that $B_n(x) \geqslant b$ for all $n \geqslant 1$ and $x \in \mathbb{R}^d$. Since α is non-decreasing it holds

$$\sup_{n \geqslant 1} \int \alpha \left(b N_n(x) \right) \mu_n(dx) \leqslant M.$$

Set $\alpha^*(t) = \sup_{s \geqslant 0} \{ st - \alpha(s) \}$, $t \geqslant 0$, and note that α^* is non-decreasing, finite valued and vanishes at 0. If $f : \mathbb{R}^d \to \mathbb{R}_+$ is a nonnegative function, then using Young's inequality $st \leqslant \alpha(s) + \alpha^*(t)$, for all $s, t \geqslant 0$ it is easily seen that for all $u > 0$, it holds

$$\int u f \, d\eta_n = \int u f N_n \, d\mu_n \leqslant \frac{1}{b} \int \alpha^*(uf) \, d\mu_n + \frac{1}{b} \int \alpha(b N_n) \, d\mu_n$$

and so

$$\int f \, d\eta_n \leqslant \frac{1}{bu} \int \alpha^*(uf) \, d\mu_n + \frac{M}{bu}. \tag{40}$$

In particular, if $f = \mathbf{1}_A$ with A a measurable set, then it holds

$$\eta_n(A) \leqslant \frac{\max(M;1)}{b} \psi(\mu_n(A)),$$

where

$$\psi(t) = \inf_{u > 0} \left\{ \frac{\alpha^*(u)}{u} t + \frac{1}{u} \right\}, \qquad t \geqslant 0.$$

Since the sequence $(\mu_n)_{n \geqslant 1}$ converges to μ, it is tight and so for all $\varepsilon > 0$, there exists a compact set K_ε such that $\sup_{n \geqslant 1} \mu_n(\mathbb{R}^d \backslash K_\varepsilon) \leqslant \varepsilon$. Choosing $A = \mathbb{R}^d \backslash K_\varepsilon$ one thus sees that

$$\sup_{n \geqslant 0} \eta_n(\mathbb{R}^d \backslash K_\varepsilon) \leqslant \frac{\max(M;1)}{b} \psi(\varepsilon).$$

It is not difficult to check that $\psi\left(\frac{1}{\alpha(s)}\right) = \frac{s}{\alpha(s)}$, as soon as $\alpha(s) > 0$. This implies in particular that $\psi(t) \to 0$ as $t \to 0$. So the bound above shows that the sequence $(\eta_n)_{n \geqslant 1}$ is tight. Therefore, according to Prokhorov's theorem, one can extract from the sequence $(\eta_n)_{n \geqslant 1}$ a subsequence converging to some probability measure η on $\mathbb{R}^d$. For notational convenience this subsequence will still be denoted by $(\eta_n)_{n \geqslant 1}$. Letting $n \to \infty$ in (40), one sees that for all bounded continuous and nonnegative function f, it holds

$$\int f \, d\eta \leqslant \frac{1}{bu} \int \alpha^*(uf) \, d\mu + \frac{M}{bu}.$$

If A is a compact subset of $\mathbb{R}^d$, then considering the sequence $f_k(x) = [1 - kd(x,A)]_+$, $k \geqslant 0$, of bounded continuous nonnegative functions which converges monotonically to $\mathbf{1}_A$, one sees that

$$\eta(A) \leqslant \frac{1}{bu}\alpha^*(u)\mu(A) + \frac{M}{bu}, \qquad \forall u > 0.$$

Since both η and μ are inner regular, this inequality is actually true for any measurable set A of $\mathbb{R}^d$. In particular, if $\mu(A) = 0$ then it easily follows that $\eta(A) = 0$ and so η is absolutely continuous with respect to μ. Since η_n converges, the sequence $\pi_n \in \Pi(\eta_n, \nu)$ is also tight. One can thus assume without loss of generality that it converges to some $\pi \in \Pi(\eta, \nu)$ in the weak topology. If u is some compactly supported function on $\mathbb{R}^d$, it holds

$$\iint u(x)y\,\pi_n(dxdy) = \int u(x)\int y\,q_n^x(dy)\mu_n(dx) = \int u(x)x\,\mu_n(dx)$$

and letting $n \to \infty$ gives

$$\iint u(x)y\,\pi(dxdy) = \int u(x)x\,\mu(dx)$$

which reads, writing $\pi(dxdy) = N(x)\mu(dx)p^x(dy)$,

$$\int u(x)\left(N(x)\int y\,p^x(dy) - x\right)\mu(dx) = 0.$$

Since this holds for all compactly supported continuous functions u, one concludes that for μ almost all $x \in \mathbb{R}^d$

$$\int y\,q^x(dy) = x$$

with $q^x = N(x)p^x$, which completes the proof. □

5.4. **Study of a particular class of nonpositive conical transport problems.** In this section, we consider a *nonpositive* cost function

$$F : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}_-$$

such that $F(x, \lambda z)/\lambda \to 0$ as $\lambda \to +\infty$ for all $x \in \mathcal{X}$, $z \in \mathcal{Z}\backslash\{0\}$. The following result shows in particular that under mild additional assumptions on F any weak solution is strong.

Theorem 5.4. *Assume that*

- 0 *does not belong to* $\mathrm{co}(\mathcal{Y})$,
- F *is nonpositive, satisfies Assumption* (A') *and is continuous on* $\mathcal{X} \times \mathcal{Z}$,
- *For all* $x \in \mathcal{X}$ *and* $y \in \mathcal{Y}$, *it holds*
$$\frac{F(x,\lambda y)}{\lambda} \to 0$$
as $\lambda \to +\infty$.
- *It holds*
$$\sup_{x\in\mathcal{X}, y\in\mathcal{Y}} F(x, \lambda y) \to -\infty$$
as $\lambda \to +\infty$.

Then it holds

$$\mathcal{I}_c(\mu,\nu) = \sup_{\varphi\in\Phi^+(\mathcal{Z})\cap L^1(\nu)} \left\{\int Q_F\varphi(x)\,\mu(dx) - \int \varphi(y)\,\nu(dy)\right\}, \qquad \forall \nu \in \mathcal{P}(\mathcal{Y}) \tag{41}$$

where $\Phi^+(\mathcal{Z})$ *is the set of all nonnegative, lower semicontinuous, convex and positively* 1*-homogenous functions* $\varphi : \mathcal{Z} \to \mathbb{R}_+ \cup \{+\infty\}$ *and where*

$$Q_F\varphi(x) = \inf_{z\in\mathcal{Z}} \{\varphi(z) + F(x,z)\}, \qquad x \in \mathcal{X}.$$

Furthermore, the supremum in (41) is attained at some function $\bar\varphi \in \Phi^+(\mathcal{Z}) \cap L^1(\nu)$, which is positive on $\mathcal{Z}\backslash\{0\}$. Finally, any weak solution for the transport problem between μ and ν is a strong solution.

The assumptions of Theorem 5.4 are for instance satisfied by $F : \mathcal{X} \times \mathbb{R}_+^d \to (-\infty, 0]$ of the following form

$$F(x, z) = -\|A(x)z\|_\sigma^\eta, \qquad x \in \mathcal{X}, z \in \mathbb{R}_+^d, \tag{42}$$

where $0 < \eta < 1$, $A : \mathcal{X} \to M_{>0}(\mathbb{R}^d)$ a continuous function taking values in the space $M_{>0}(\mathbb{R}^d)$ of $d \times d$ matrices with positive entries, and for $0 < \sigma \leqslant 1$, the σ-"norm" is defined by

$$\|z\|_\sigma = \left(\sum_{i=1}^d |z_i|^\sigma\right)^{1/\sigma}, \qquad z \in \mathbb{R}^d.$$

It is well known that $\|\cdot\|_\sigma$ satisfies the following reverse triangle inequality on $\mathbb{R}_+^d$:

$$\|z_1 + z_2\|_\sigma \geqslant \|z_1\|_\sigma + \|z_2\|_\sigma, \qquad \forall z_1, z_2 \in \mathbb{R}_+^d.$$

This easily implies that the function F defined by (42) is convex with respect to its second variable. Since $0 < \eta < 1$, it is also clear that for every $x \in \mathcal{X}$ and $z \in \mathbb{R}_+^d$, $F(x, \lambda z)/\lambda \to 0$ as $\lambda \to +\infty$. Finally, if $\mathcal{Y}$ is a compact subset included in $(0,\infty)^d$, then it is easy to see that $\sup_{x\in\mathcal{X}, y\in\mathcal{Y}} F(x, \lambda y) \to -\infty$ as $\lambda \to +\infty$.

Remark 5.2. *Cost functions of the form (42) are considered in [15] to represent minus the output of a firm x when it hires a worker of type z. In this context, the variable φ appearing in the dual formulation of $\mathcal{I}_c(\mu, \nu)$ corresponds to a wage function and is thus naturally nonnegative.*

Proof. Applying Theorem 5.1 to the cost function c yields to

$$\mathcal{I}_c(\mu, \nu) = \sup_{\varphi \in \Phi(\mathcal{Z}) \cap L^1(\nu)} \left\{ \int Q_F\varphi(x)\, \mu(dx) - \int \varphi(y)\, \nu(dy) \right\}, \qquad \forall \nu \in \mathcal{P}(\mathcal{Y}).$$

Let us show that the supremum can be restricted to nonnegative functions. Indeed, let $y_0 \in \mathrm{co}(\mathcal{Y})$ be such that $\varphi(y_0) = \inf_{\mathrm{co}(\mathcal{Y})} \varphi$ and assume that $\varphi(y_0) < 0$. Since $F \leqslant 0$, it holds for all $\lambda \geqslant 0$

$$Q_F\varphi(x) \leqslant \inf_{\lambda > 0} \varphi(\lambda y_0) = \inf_{\lambda > 0} \lambda \varphi(y_0) = -\infty.$$

Therefore, such functions φ can be dropped from the supremum. According to Theorem 5.1, we also know that the supremum in (41) is reached at some $\bar\varphi \in \Phi^+(\mathcal{Z}) \cap L^1(\nu)$. Let us show that this function φ is positive over $\mathcal{Z}\backslash\{0\}$. Consider again $y_0 \in \mathrm{co}(\mathcal{Y})$ such that $\varphi(y_0) = \inf_{\mathrm{co}(\mathcal{Y})} \varphi$ and set $a = \varphi(y_0) \geqslant 0$. Define, for all $u \geqslant 0$,

$$\psi(u) = \inf_{\lambda > 0} \left\{ \lambda u + \sup_{x \in \mathcal{X}, y \in \mathcal{Y}} F(x, \lambda y) \right\}.$$

Observe that

$$Q_F\varphi(x) \leqslant \inf_{\lambda > 0} \{\varphi(\lambda y_0) + F(x, \lambda y_0)\} = \inf_{\lambda > 0} \{\lambda a + F(x, \lambda y_0)\} \leqslant \psi(a),$$

where we used that since $y_0 \in \mathrm{co}(\mathcal{Y})$, it holds $F(x, \lambda y_0) \leqslant \sup_{y \in \mathcal{Y}} F(x, \lambda y)$. Therefore,

$$-\infty < \mathcal{I}_c(\mu, \nu) = \int Q_F\bar\varphi(x)\, \mu(dx) - \int \bar\varphi(y)\, \nu(dy) \leqslant \psi(a) - a \leqslant \psi(a).$$

By assumption $\psi(0) = -\infty$, so $a > 0$. Therefore $\bar\varphi$ is positive on $\mathrm{co}(\mathcal{Y})$. Since $\mathcal{Z} = \mathbb{R}_+\mathrm{co}(\mathcal{Y})$, we conclude that $\bar\varphi$ is positive on $\mathcal{Z}\backslash\{0\}$.

Now let us show that the transport problem admits only strong solutions. According to Theorem 2.1 (applied to the cost function c which satisfies condition (C)), we know there exists $\bar{\pi} \in \Pi(\mathrm{Supp}(\mu), \nu)$ such that

$$\mathcal{I}_c(\mu,\nu) = \overline{\mathcal{I}}_c(\mu,\nu) = \bar{I}_c^{\mu}[\bar{\pi}], \tag{43}$$

where, for all $\pi \in \Pi(\mathrm{Supp}(\mu), \nu)$,

$$\bar{I}_c^{\mu}[\pi] = \int F\left(x, \frac{d\pi_1^{ac}}{d\mu}(x) \int y\, \pi^x(dy)\right) \mu(dx)$$

(because, by assumption, $F'_\infty(x,z) = 0$ for all x, z). Denoting by $S(x) = \frac{d\bar{\pi}_1^{ac}}{d\mu}(x) \int y\, \bar{\pi}^x(dy)$, it holds

$$\begin{aligned}
\mathcal{I}_c(\mu,\nu) &= \int Q_F\bar{\varphi}(x)\,\mu(dx) - \int \bar{\varphi}(y)\,\nu(dy)\\
&\leqslant \int \bar{\varphi}(S(x)) + F(x,S(x))\,\mu(dx) - \int \bar{\varphi}(y)\,\nu(dy)\\
&\overset{(i)}{\leqslant} \int \left(\int \bar{\varphi}(y)\,\bar{\pi}^x(dy)\right) \bar{\pi}_1^{ac}(dx) + \int F(x,S(x))\,\mu(dx) - \int \bar{\varphi}(y)\,\nu(dy)\\
&\overset{(ii)}{=} \bar{I}_c^{\mu}[\bar{\pi}] - \int \left(\int \bar{\varphi}(y)\,\bar{\pi}^x(dy)\right) \bar{\pi}_1^{s}(dx)\\
&\overset{(iii)}{=} \mathcal{I}_c(\mu,\nu) - \int \left(\int \bar{\varphi}(y)\,\bar{\pi}^x(dy)\right) \bar{\pi}_1^{s}(dx),
\end{aligned}$$

where

- (i) comes from the positive 1-homogeneity of $\bar{\varphi}$ and Jensen's inequality,
- (ii) comes from the definition of $\bar{I}_c^{\mu}[\bar{\pi}]$ and the fact that

$$\int \bar{\varphi}(y)\,\nu(dy) = \int \left(\int \bar{\varphi}(y)\,\bar{\pi}^x(dy)\right) \bar{\pi}_1(dx) = \int \left(\int \bar{\varphi}(y)\,\bar{\pi}^x(dy)\right) \bar{\pi}_1^{ac}(dx) + \int \left(\int \bar{\varphi}(y)\,\bar{\pi}^x(dy)\right) \bar{\pi}_1^{s}(dx),$$

- and (iii) comes from (43).

We conclude that

$$\int \left(\int \bar{\varphi}(y)\,\bar{\pi}^x(dy)\right) \bar{\pi}_1^{s}(dx) \leqslant 0.$$

Note that $\int \bar{\varphi}(y)\,\bar{\pi}^x(dy) \geqslant a > 0$, so the only possibility is that $\bar{\pi}_1^s = 0$ (no singular part). Therefore, $\bar{\pi} \in \Pi(\ll \mu, \nu)$ and $\bar{I}_c^{\mu}[\bar{\pi}] = I_c^{\mu}[\bar{\pi}] = \mathcal{I}_c(\mu,\nu)$ and so $\bar{\pi}$ is a strong solution. □

5.5. **Structure of solutions for conical cost functions.** In all the subsection, we assume that F is a function satisfying Assumption (A') and that c is the conical cost function associated to F and defined by (33).

The following result gives an interpretation of the transport cost $\mathcal{I}_c(\mu,\nu)$ as a shortest transport distance between μ and the set of probability measures dominated by ν in the order $\leqslant_{phc}$.

Theorem 5.5. *Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ be such that $\mathcal{I}_c(\mu,\nu) < +\infty$, and assume that the convex hull of the support of ν does not contain 0. Then the following identity holds*

$$\mathcal{I}_c(\mu,\nu) = \inf_{\gamma \leqslant_{phc} \nu} \mathcal{T}_F(\mu,\gamma), \tag{44}$$

where and $\mathcal{T}_F$ denotes the classical transport cost associated to the cost function F:

$$\mathcal{T}_F(\mu,\gamma) = \inf_{\pi \in \Pi(\mu,\gamma)} \iint F(x,z)\,\pi(dxdz), \qquad \forall \mu \in \mathcal{P}(\mathcal{X}), \forall \gamma \in \mathcal{P}(\mathcal{Z}).$$

Moreover, suppose that $\bar{q}$ is a nonnegative kernel solution to the transport problem (10)*, consider the map $\bar{S}$ defined by*

$$\bar{S}(x) = \int y\, \bar{q}^x(dy), \qquad x \in \mathcal{X},$$

and denote by $\bar{\nu}$ the image of μ under the map $\bar{S}$. Then the following holds :

- *the probability measure $\bar{\nu}$ is dominated by ν in the positively 1-homogenous convex order,*
- *it holds*

$$\mathcal{I}_c(\mu,\nu) = \int F(x, \bar{S}(x))\, \mu(dx) = \inf_{\gamma \leqslant_{phc} \nu} \mathcal{T}_F(\mu,\gamma)$$

Note that the map $\bar{S}$ provides an optimal transport map between μ and $\bar{\nu}$ for the cost $\mathcal{T}_F$. The proof below adapts the proof of [20, Proposition 1.1] and [4, Lemma 6.1] to our setting.

Proof. Let q be a nonnegative kernel such that $\mu q = \nu$. Thanks to Jensen's inequality, for all positively 1-homogenous convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$, it holds

$$\int \varphi(S(x))\, \mu(dx) \leqslant \iint \varphi(y)\, q^x(dy)\, \mu(dx) = \int \varphi(y)\, \nu(dy),$$

where $S(x) = \int y\, q^x(dy)$, $x \in \mathcal{X}$, and so $S_\# \mu$ is dominated by ν in the positively homogenous convex order. Therefore,

$$\int F(x, S(x))\, \mu(dx) \geqslant \inf_{\gamma \leqslant_{phc} \nu} \inf_{\pi \in \Pi(\mu,\gamma)} \int F(x,z)\, \pi(dxdz).$$

Optimizing over q shows that

$$\inf_{\gamma \leqslant_{phc} \nu} \mathcal{T}_F(\mu,\gamma) \leqslant \mathcal{I}_c(\mu,\nu).$$

Let us prove the converse inequality. Let $\gamma \leqslant_{phc} \nu$ and $\pi \in \Pi(\mu,\gamma)$. Since $\gamma \leqslant_{phc} \nu$, Theorem 5.2 shows that there exists a nonnegative kernel $(r^z)_{z \in \mathbb{R}^d}$ such that $\int r^z(dy)\, \gamma(dz) = \nu(dy)$ and $\int y\, r^z(dy) = z$ for γ almost all z. Write

$$\pi(dxdz) = \mu(dx) p^x(dz)$$

where p is a probability kernel, and consider the nonnegative kernel q defined by

$$q^x(dy) = \int p^x(dz) r^z(dy),$$

which satisfies $\mu q = \nu$. Moreover, for μ almost all x, it holds

$$\int y q^x(dy) = \int p^x(dz) \int y r^z(dy) = \int z p^x(dz).$$

Thus,

$$\begin{aligned}\int F(x,z)\, \pi(dxdz) = \iint F(x,z)\, \mu(dx) p^x(dz) &\geqslant \int F\left(x, \int z\, p^x(dz)\right) \mu(dx) \\ &= \int F\left(x, \int y q^x(dy)\right) \mu(dx) \geqslant \mathcal{I}_c(\mu,\nu).\end{aligned}$$

Optimizing over π and over $\gamma \leqslant_{phc} \nu$ gives that

$$\inf_{\gamma \leqslant_{phc} \nu} \mathcal{T}_F(\mu,\gamma) \geqslant \mathcal{I}_c(\mu,\nu)$$

which proves (44). Now if $\bar{q}$ is a strong solution, then $\bar{\nu} = \bar{S}_\# \mu \leqslant_{phc} \nu$ and

$$\mathcal{I}_c(\mu,\nu) = \int F(x, \bar{S}(x))\, \mu(dx),$$

which completes the proof. □

The following result considers the particular case of dimension 1.

Proposition 5.1. *Let $\nu \in \mathcal{P}(\mathbb{R}_+)$ be a compactly supported probability measure with $m = \int y\,\nu(dy)$ and denote by*

$$C_m = \left\{\gamma \in \mathcal{P}_1(\mathbb{R}_+) : \int x\,\gamma(dx) = m\right\}.$$

Then it holds

- $C_m = \{\gamma \in \mathcal{P}_1(\mathbb{R}) : \gamma \leqslant_{phc} \nu\}$,
- *if $\mu \in \mathcal{P}(\mathcal{X})$ is such that*
$$\mathcal{I}_c(\mu,\nu) = \min_{\gamma \in C_m} \mathcal{T}_F(\mu,\gamma),$$
then, there exists a map $S_m : \mathcal{X} \to \mathbb{R}_+$ transporting μ onto $\gamma_m \in C_m$ such that
$$\mathcal{I}_c(\mu,\nu) = \mathcal{T}_F(\mu,\gamma_m) = \int F(x, S_m(x))\,\mu(dx),$$
- *the nonnegative kernel $\bar{q}$ defined by*
$$\bar{q}^x(dy) = \frac{S_m(x)}{m}\,\nu(dy), \qquad x \in \mathcal{X},$$
is a strong solution of the transport problem (10).

This result tell us that, in dimension one, once the solutions for the transport problem between μ and Dirac masses δ_m, $m \geqslant 0$, are known, then optimal solutions can be deduced for general ν on $\mathbb{R}_+$.

Proof. According to Remark 5.1, we know that for any compactly supported probability measure ν on $\mathbb{R}_+$,

$$\{\gamma \in \mathcal{P}_1(\mathbb{R}) : \gamma \leqslant_{phc} \nu\} = \left\{\gamma \in \mathcal{P}(\mathbb{R}_+) : \int x\,\gamma(dx) = \int x\,\nu(dx)\right\} = C_m.$$

According to Theorem 5.5, it holds

$$\mathcal{I}_c(\mu,\nu) = \inf_{\gamma \leqslant_{phc} \nu} \mathcal{T}_F(\mu,\gamma) = \inf_{\gamma \in C_m} \mathcal{T}_F(\mu,\gamma).$$

By assumption, this last infimum is reached at some point $\gamma'_m \in C_m$. Let $\pi \in \Pi(\mu, \gamma'_m)$ be an optimal coupling for $\mathcal{T}_F(\mu, \gamma'_m)$ and write $\pi(dxdy) = \mu(dx)p^x(dy)$. By Jensen's inequality it holds

$$\mathcal{T}_F(\mu,\gamma'_m) = \iint F(x,y)p^x(dy)\mu(dx) \geqslant \int F(x, S_m(x))\,\mu(dx),$$

where $S_m(x) = \int y p^x(dy)$. Denoting by $\gamma_m = (S_m)_\#\mu$, one sees that $\gamma_m \leqslant_c \gamma'_m$ and in particular, $\gamma_m \in C_m$. Therefore, one gets

$$\inf_{\gamma \in C_m} \mathcal{T}_F(\mu,\gamma) = \mathcal{T}_F(\mu,\gamma'_m) \geqslant \int F(x, S_m(x))\,\mu(dx) \geqslant \mathcal{T}_F(\mu,\gamma_m) \geqslant \inf_{\gamma \in C_m} \mathcal{T}_F(\mu,\gamma).$$

This proves that γ'_m can be replaced by γ_m and that S_m is an optimal transport map (for the cost F) between μ and γ_m. The nonnegative kernel $\bar{q}$ defined in Proposition 5.1 satisfies

$$\int f(y)\,\bar{q}^x(dy)\mu(dx) = \frac{1}{m}\iint f(y)S_n(x)\,\mu(dx)\nu(dy) = \int f(y)\,\nu(dy)\frac{\int y\,\gamma_m(dy)}{m} = \int f(y)\,\nu(dy)$$

and so $\mu\bar{q} = \nu$. Moreover, for all x, $\int y\,\bar{q}^x(dy) = S_m(x)$ and so

$$\int F(x, S_m(x))\,\mu(dx) = \int F\left(x, \int y\,\bar{q}^x(dy)\right)\,\mu(dx),$$

which shows that $\bar{q}$ is a strong solution. □

The next result establishes a variant of Theorem 5.5 involving the classical convex order $\leqslant_c$.

Theorem 5.6. *Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$, and assume that $\bar{\pi} \in \Pi(\ll \mu, \nu)$ is a strong solution to the transport problem (10) for the conical cost function c defined by (33). Let $\bar{\eta}(dx) = \bar{N}(x)\,\mu(dx)$ be the first marginal of $\bar{\pi}$ and consider the map $\bar{T}$ defined by*

$$\bar{T}(x) = \int y\,\bar{\pi}^x(dy), \qquad x \in \mathcal{X},$$

and denote by $\tilde{\nu}$ the image of $\bar{\eta}$ under the map $\bar{T}$. Consider the function $G : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ defined for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ by

$$G(x, y) = \frac{F(x, \bar{N}(x)y)}{\bar{N}(x)},$$

if $\bar{N}(x) > 0$, and $G(x, y) = 0$ (or any other arbitrary value) otherwise. The following holds :

- *the probability measure $\tilde{\nu}$ is dominated by ν in the convex order,*
- *it holds*
$$\mathcal{I}_c(\mu, \nu) = \int G(x, \bar{T}(x))\,\bar{\eta}(dx) = \inf_{\gamma \leqslant_c \nu} \mathcal{T}_G(\bar{\eta}, \gamma),$$
where we denote by $\mathcal{T}_G$ the Monge-Kantorovich optimal transport cost associated to the cost function G:
$$\mathcal{T}_G(\eta, \gamma) = \inf_{\pi \in \Pi(\eta,\gamma)} \iint G(x, z)\,\pi(dxdz), \qquad \forall \eta \in \mathcal{P}(\mathcal{X}), \forall \gamma \in \mathcal{P}(\mathcal{Y}).$$

In other words, the probability measure $\tilde{\nu}$ turns out to be the closest point to $\bar{\eta}$ among the set $\{\gamma \in \mathcal{P}(\mathcal{Y}) : \gamma \leqslant_c \nu\}$ for the transport "distance" $\mathcal{T}_G$. Moreover, the map $\bar{T}$ (which is sometimes called the barycentric projection of the coupling $\bar{\pi}$) provides an optimal transport map between $\bar{\eta}$ and $\tilde{\nu}$ for the cost $\mathcal{T}_G$.

Remark 5.3. *Suppose that $\bar{\pi}$ is a strong solution to the transport problem 10, denote by $\bar{\eta} = \bar{N}\,\mu$ its first marginal, and consider the nonnegative kernel $\bar{q}^x = \bar{N}(x)\pi^x$, $x \in \mathcal{X}$. Then for all $x \in \mathcal{X}$, $\bar{S}(x) = N(x)\bar{T}(x)$. Note however that $\bar{\nu}$ and $\tilde{\nu}$ are in general two different probability measures, so that the conclusions of Theorems 5.5 and 5.6 are not equivalent. Nevertheless, for all positively homogenous function φ it holds $\int \varphi(y)\,\bar{\nu}(dy) = \int \varphi(y)\,\tilde{\nu}(dy)$.*

Proof. Thanks to Jensen's inequality, for all convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$, it holds

$$\int \varphi(\bar{T}(x))\,\bar{\eta}(dx) \leqslant \iint \varphi(y)\,\bar{\pi}^x(dy)\,\bar{\eta}(dx) = \int \varphi(y)\,\nu(dy),$$

and so $\tilde{\nu} = \bar{T}_\#\bar{\eta}$ is dominated by ν in the convex order. Therefore,

$$\mathcal{I}_c(\mu, \nu) = \int G(x, \bar{T}(x))\,\bar{\eta}(dx) \geqslant \inf_{\gamma \leqslant_c \nu} \inf_{\pi \in \Pi(\bar{\eta},\gamma)} \int G(x, z)\,\pi(dxdz).$$

Let us prove the converse inequality. Let $\gamma \leqslant_c \nu$ and $\pi \in \Pi(\bar{\eta}, \gamma)$. Since $\gamma \leqslant_c \nu$ there exists a martingale coupling m between γ and ν, that is to say a probability measure $m \in \Pi(\gamma, \nu)$ such that $m(dzdy) = \gamma(dz)m^z(dy)$ and $\int y\,m^z(dy) = z$ for γ almost every z. Write

$$\pi(dxdz) = \bar{\eta}(dx)p^x(dz) = \gamma(dz)r^z(dx)$$

and consider the coupling $\bar{\pi}$ defined by

$$\bar{\pi}(dxdy) = \int r^z(dx)m^z(dy)\gamma(dz).$$

It is easily seen that $\bar{\pi} \in \Pi(\bar{\eta}, \nu)$. Also,

$$\bar{\pi}(dxdy) = \bar{\eta}(dx) \int p^x(dz)m^z(dy)$$

and so $\bar{\pi}^x(dy) = \int p^x(dz) m^z(dy)$. Moreover it holds

$$\int y \bar{\pi}^x(dy) = \int y \int p^x(dz) m^z(dy) = \int z\, p^x(dz).$$

Thus,

$$\begin{aligned}\iint G(x,z)\,\pi(dxdz) &= \iint G(x,z)\,\bar{\eta}(dx) p^x(dz) \geqslant \int G\left(x, \int z\, p^x(dz)\right)\,\bar{\eta}(dx) \\ &= \int G\left(x, \int y\, \bar{\pi}^x(dy)\right)\,\bar{\eta}(dx) = I_c^\mu[\bar{\pi}] \geqslant \mathcal{I}_c(\mu,\nu)\end{aligned}$$

which completes the proof. □

The aim of the next result is to understand the articulation between primal and dual optimizers.

Theorem 5.7. *Let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ be such that $\mathcal{I}_c(\mu,\nu) < +\infty$ and assume that $\bar{q}$ is a kernel solution and $\bar{\varphi} \in \Phi(\mathcal{Z}) \cap L^1(\nu)$ a dual optimizer:*

$$\mathcal{I}_c(\mu,\nu) = \int F\left(x, \int y\, \bar{q}^x(dy)\right)\,\mu(dx) = \int Q_F \bar{\varphi}(x)\,\mu(dx) - \int \bar{\varphi}(y)\,\nu(dy).$$

Define $\bar{S}(x) = \int y\, \bar{q}^x(dy)$, $x \in \mathcal{X}$.

(1) *For μ almost every $x \in \mathcal{X}$, it holds*

$$Q_F \bar{\varphi}(x) = \bar{\varphi}(\bar{S}(x)) + F(x, \bar{S}(x)).$$

(2) *If M denotes the set of $x \in \mathcal{X}$ for which the support $K(x) \subset \mathcal{Y}$ of $\bar{q}^x$ contains at least two points, then for $\bar{\eta}$ almost $x \in M$, the function $\bar{\varphi}$ is affine on the convex hull of $K(x)$: there exist $u_x \in \mathbb{R}^d$ and $v_x \in \mathbb{R}$ such that $\bar{\varphi}(z) = u_x \cdot z + v_x$ for all $z \in \mathrm{co}(K(x))$.*

(3) *If F is strictly convex with respect to its second variable, then the map $\bar{S}(x) = \int y\, \bar{q}^x(dy)$, $x \in \mathcal{X}$, is μ-almost surely unique among all strong solutions $\bar{q}$ of the transport problem.*

(4) *If F is strictly convex with respect to its second variable and if for all $x \in \mathcal{X}$ there exist $A_x \in \mathbb{R}$ and $M_x > 0$ such that $\bar{\varphi}(z) + F(x,z) \geqslant A_x + M_x\|z\|$ for all $z \in \mathcal{Z}$, then for all $x \in \mathcal{X}$ the map $\bar{\varphi}^* \,\square\, F^*(x,\,\cdot\,)$ is differentiable in a neighborhood of 0 and it holds*

$$\bar{S}(x) = \nabla \left(\bar{\varphi}^* \,\square\, F^*(x,\,\cdot\,)\right)(0)$$

for μ almost all x.

In the result above we denoted

$$\bar{\varphi}^*(u) = \sup_{z \in \mathcal{Z}} \{z \cdot u - \bar{\varphi}(z)\}, \qquad u \in \mathbb{R}^d,$$

and for $x \in \mathcal{X}$

$$F^*(x,u) = \sup_{z \in \mathcal{Z}} \{z \cdot u - F(x,z)\}, \qquad u \in \mathbb{R}^d,$$

the Fenchel-Legendre transforms of the functions $\bar{\varphi}$ and $F(x,\,\cdot\,)$ extended by $+\infty$ outside $\mathcal{Z}$. Moreover, we recall that the infimum convolution between $\bar{\varphi}^*$ and $F^*(x,\,\cdot\,)$ is defined by

$$\bar{\varphi}^* \,\square\, F^*(x,\,\cdot\,)(u) = \inf_{v \in \mathbb{R}^d} \{\bar{\varphi}^*(v) + F^*(x, u-v)\}, \qquad u \in \mathbb{R}^d.$$

Remark 5.4. *Denoting by $C(\bar{\varphi}) = \{u \in \mathbb{R}^d : u \cdot z \leqslant \bar{\varphi}(z), \forall z \in \mathcal{Z}\}$, it is easily seen that*

$$\bar{\varphi}^* = \chi_{C(\bar{\varphi})},$$

where $\chi_{C(\bar{\varphi})}(u) = 0$ if $u \in C(\bar{\varphi})$ and $+\infty$. So, it holds

$$\bar{\varphi}^* \,\square\, F^*(x,\,\cdot\,)(u) = \inf_{v \in C(\bar{\varphi})} \{F^*(x, u-v)\}, \qquad u \in \mathbb{R}^d. \tag{45}$$

Proof. By optimality of $\bar\varphi$ and $\bar q$, it holds

$$\begin{aligned}\mathcal{I}_c(\mu,\nu) &= \int Q_F\bar\varphi(x)\,\mu(dx) - \int \bar\varphi(y)\,\nu(dy)\\ &\leqslant \int \bar\varphi(\bar S(x)) + F(x,\bar S(x))\,\mu(dx) - \int \bar\varphi(y)\,\nu(dy)\\ &\leqslant \int\int \bar\varphi(y)\,\bar q^x(dy) + F(x,\bar S(x))\,\mu(dx) - \int \bar\varphi(y)\,\nu(dy)\\ &= \int F(x,\bar S(x))\,\mu(dx)\\ &= \mathcal{I}_c(\mu,\nu),\end{aligned}$$

where the first inequality comes from the definition of $Q_F\bar\varphi$, the second from the fact that $\bar\varphi$ is convex and positive 1-homogenous. Analyzing the equality cases completes the proof of (1) and (2). Now, assume that F is strictly convex with respect to its second variable, and consider $\bar r$ another minimizing nonnegative kernel and define $\bar U(x) = \int y\,\bar r^x(dy)$, $x \in \mathcal{X}$. According to what precedes, for μ almost all $x \in \mathcal{X}$, the points $\bar S(x)$ and $\bar U(x)$ minimize the function

$$z \mapsto \bar\varphi(z) + F(x,z), \qquad z \in \mathcal{Z}.$$

This function being strictly convex, this implies that $\bar S(x) = \bar U(x)$ and so $\bar S = \bar U$ μ a.e., which proves (3). Let us now prove (4). Consider the function $H : \mathcal{X}\times\mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$ defined by $H(x,z) = \bar\varphi(z) + F(x,z)$, $x \in \mathcal{X}$, $z \in \mathbb{R}^d$ (with $\bar\varphi$ and $F(x,\cdot\,)$ extended by $+\infty$ outside $\mathcal{Z}$). For a given $x \in \mathcal{X}$, observe that $\bar S(x)$ minimizes the convex function $H(x,\cdot\,)$ if and only if $0 \in \partial H(x,\cdot\,)(\bar S(x))$, where $\partial H(x,\cdot\,)(z)$ denotes the subdifferential of the function $H(x,\cdot\,)$ at the point z. By the well known conjugation relation of subdifferentials, it holds

$$0 \in \partial H(x,\cdot\,)(\bar S(x)) \Leftrightarrow \bar S(x) \in \partial H^*(x,\cdot\,)(0)$$

(see e.g [24, Corollary E.1.4.4]). Moreover, since $H(x,\cdot\,)$ is a sum of two convex lower-semicontinuous functions, its Fenchel-Legendre transform is given as follows:

$$H^*(x,u) = \bar\varphi^* \,\square\, F^*(x,\cdot\,)(u), \qquad u \in \mathbb{R}^d$$

(see e.g [24, Theorem E.2.3.2] and note that $F(x,\cdot\,)$ is finite over $\mathcal{Z}$ which contains the relative interior of the domain of $\bar\varphi$). The assumed lower bound on H easily implies that, for all $x \in \mathcal{X}$, the function $H^*(x,\cdot\,)$ takes finite values in a neighborhood of 0. Since $H(x,\cdot\,)$ is strictly convex for all $x \in \mathcal{X}$, it follows from [24, Theorem E.4.1.1] that the function $H^*(x,\cdot\,)$ is continuously differentiable on the interior of its domain. In particular, it is continuously differentiable in a neighborhood of 0, and so, for all $x \in \mathcal{X}$, $\partial H^*(x,\cdot\,)(0) = \{\nabla(\bar\varphi^* \,\square\, F^*(x,\cdot\,))(0)\}$, which completes the proof. □

Corollary 5.1. *Let $F : \mathcal{X}\times\mathcal{Z} \to \mathbb{R}$ be a cost function satisfying assumption (A') and strictly convex with respect to its second variable and let $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ be such that the convex hull of the support of ν does not contain 0.*
Assume further that F satisfies

- *Assumption (B')*
  or
- *$F : \mathcal{X}\times\mathcal{Z} \to \mathbb{R}_-$ is a nonpositive function satisfying the assumptions of Theorem 5.4.*

Then, for any kernel solution $\bar q$ and any dual optimizer $\bar\varphi$ of the transport problem (3), the map $\bar\varphi^* \,\square\, F^*(x,\cdot\,)$ is differentiable in a neighborhood of 0, for all $x \in \mathcal{X}$, and it holds

$$\bar S(x) = \int y\,\bar q^x(dy) = \nabla(\bar\varphi^* \,\square\, F^*(x,\cdot\,))(0),$$

for μ almost all $x \in \mathcal{X}$.

Proof. It is well known that Assumption (B') is equivalent to the 1-coercivity of $F(x,\cdot)$, that is to say

$$\lim_{z\in\mathcal{Z},\|z\|\to+\infty}\frac{F(x,z)}{\|z\|}\to+\infty.$$

Therefore, since the convex function $\bar\varphi$ admits at least one affine minorant, it is easily seen that for every $x\in\mathcal{X}$ there exist $A_x\in\mathbb{R}$ and $M_x>0$ such that $\bar\varphi(z)+F(x,z)\geqslant A_x+M_x\|z\|$ for all $z\in\mathcal{Z}$. We conclude using Item (4) of Theorem 5.7. Similarly, if F satisfies the assumptions of Theorem 5.4, then for all $x\in\mathcal{X}$ and $z\in\mathcal{Z}$ it holds $F(x,\lambda z)/\lambda\to 0$ as $\lambda\to+\infty$. This is actually equivalent to the fact that

$$\lim_{z\in\mathcal{Z},\|z\|\to+\infty}\frac{F(x,z)}{\|z\|}=0.$$

Let us briefly sketch the proof. Let $z_k\in\mathcal{Z}$, $k\geqslant 0$, be a sequence such that $\lambda_k=\|z_k\|\to+\infty$ monotonically, as $k\to+\infty$. Define $u_k=z_k/\lambda_k$, $k\geqslant 0$. By compactness, one can assume without loss of generality that $u_k\to u\in\mathcal{Z}$ as $k\to+\infty$. Then, by convexity it holds, for all $k\geqslant n$

$$\frac{F(x,z_k)-F(x,0)}{\lambda_k}=\frac{F(x,\lambda_k u_k)-F(x,0)}{\lambda_k}\geqslant\frac{F(x,\lambda_n u_k)-F(x,0)}{\lambda_n}$$

Thus letting $k\to+\infty$, one gets

$$\liminf_{k\to\infty}\frac{F(x,z_k)}{\|z_k\|}\geqslant\frac{F(x,\lambda_n u)-F(x,0)}{\lambda_n}$$

and letting $n\to+\infty$ gives that

$$\liminf_{k\to\infty}\frac{F(x,z_k)}{\|z_k\|}\geqslant F'_\infty(x,u)=0.$$

Since F is nonpositive this proves the claim. Now, if $\bar\varphi$ is some dual optimizer, we know by Theorem 5.4 that $\bar\varphi>0$ on $\mathcal{Z}\backslash\{0\}$. Thus denoting by $M=\inf_{\|u\|=1,u\in\mathcal{Z}}\bar\varphi(u)>0$, one sees that $\bar\varphi(z)\geqslant M\|z\|$, for all $z\in\mathcal{Z}$. And so

$$\limsup_{\|z\|\to+\infty}\frac{\bar\varphi(z)+F(x,z)}{\|z\|}\geqslant M>0,$$

and we conclude using Item (4) of Theorem 5.7. □

Let us emphasize a particularly simple case related to Brenier Theorem [11, 12]. In the following result, adapted from [20, Theorem 1.2], we assume that $\mathcal{X}\subset\mathbb{R}^d$ is a compact subset of $\mathbb{R}^d$ and we consider the cost function $c_2:\mathcal{X}\times\mathcal{M}(\mathcal{Y})\to\mathbb{R}_+$ defined by

$$c_2(x,m)=\frac{1}{2}\left\|x-\int y\,m(dy)\right\|_2^2,\qquad x\in\mathcal{X},m\in\mathcal{M}(\mathcal{Y}),$$

which corresponds to the function $F_2:\mathbb{R}^d\times\mathbb{R}^d\to\mathbb{R}^+:(x,z)\mapsto\frac{1}{2}\|x-z\|_2^2$, where $\|\cdot\|_2$ is the standard Euclidean norm.

Theorem 5.8. *Let $\mu\in\mathcal{P}(\mathcal{X})$ and $\nu\in\mathcal{P}(\mathcal{Y})$ such that the convex hull of the support of ν does not contain 0. Then there exists a unique probability measure $\bar\nu\in\mathcal{P}(\mathcal{Z})$ such that*

$$\mathcal{I}_{c_2}(\mu,\nu)=\frac{1}{2}\inf_{\eta\leqslant_{phc}\nu}W_2^2(\mu,\eta)=\frac{1}{2}W_2^2(\mu,\bar\nu).\tag{46}$$

Moreover, there exists a closed convex set $C\subset\mathbb{R}^d$ such that for any nonnegative kernel $\bar q$ minimizing $\mathcal{I}_c(\mu,\nu)$, it holds

$$\bar S(x)=\int y\,\bar q^x(dy)=x-p_C(x),$$

for μ almost every x, where $p_C:\mathbb{R}^d\to\mathbb{R}^d$ is the orthogonal projection onto C. The probability $\bar\nu$ is the image of μ under the map $x\mapsto x-p_C(x)$.

Proof. The cost function c_2 clearly satisfies Assumption (B) and so, according to Theorem 2.2, the transport problem (3) between μ and ν admits kernel solutions. According to Theorem 5.1, it also admits dual optimizers. Let $\bar{\varphi} \in \Phi(\mathcal{Z}) \cap L^1(\nu)$ be a dual optimizer (extended by $+\infty$ outside $\mathcal{Z}$). As observed in Remark 5.4, $\bar{\varphi}^* = \chi_C$ for some closed convex set C, and according to (45) it holds

$$\bar{\varphi}^* \,\square\, F_2(x,\cdot)(u) = \inf_{v\in C} F_2^*(x, u-v) = \inf_{v\in C}\left\{\frac{1}{2}\|u-v\|_2^2 + (u-v)\cdot x\right\} = -\frac{\|x\|_2^2}{2} + \frac{1}{2}d_C^2(x+u),$$

where $d_C(a) = \inf_{v\in C}\|a-v\|_2$. It is well known that d_C^2 is differentiable over $\mathbb{R}^d$ and that

$$\nabla\left(\frac{1}{2}d_C^2\right)(a) = a - p_C(a), \qquad \forall a \in \mathbb{R}^d.$$

Therefore,

$$\nabla\left(\bar{\varphi}^* \,\square\, F_2(x,\cdot)\right)(0) = x - p_C(x), \qquad \forall x \in \mathcal{X}.$$

According to Corollary 5.1, for any kernel solution $\bar{q}$, it holds

$$\int y\,\bar{q}^x(dy) = x - p_C(x),$$

for μ almost all $x \in \mathbb{R}^d$. According to Theorem 5.5, we conclude that the probability measure $\bar{\nu}$ defined as the push forward of μ under $x \mapsto x - p_C(x)$ satisfies (46). The uniqueness of $\bar{\nu}$ is obtained as in [20, Proposition 1.1]. □

Remark 5.5. *Let us give a geometric justification of the fact that $x - p_C(x)$ belongs to $\mathcal{Z}$ for all $x \in \mathbb{R}^d$. Denoting by $\bar{\varphi}$ the dual optimizer used in the proof (extended by $+\infty$ outside $\mathcal{Z}$), one has $\mathcal{Z}_{\bar{\varphi}} := \operatorname{cl}\operatorname{dom}(\bar{\varphi}) \subset \mathcal{Z}$. But, since $\bar{\varphi}$ is the support function of $C = \{u : u\cdot x \leqslant \bar{\varphi}(x), \forall x \in \mathcal{Z}\}$, one has according to* [24, Proposition C.2.2.4]

$$C_\infty^\circ = \mathcal{Z}_{\bar{\varphi}},$$

where

- *C_∞ denotes the asymptotic cone of C, defined by*
$$C_\infty = \bigcap_{t>0}\frac{C - x_o}{t},$$
with $x_o \in C$ some arbitrary point,
- *C_∞° denotes the polar cone of C_∞ defined by*
$$C_\infty^\circ = \{x \in \mathbb{R}^d : x\cdot y \leqslant 0, \forall y \in C_\infty\}.$$

By definition of the orthogonal projection on C, it holds

$$(x - p_C(x))\cdot(a - p_C(x)) \leqslant 0, \qquad \forall x \in \mathbb{R}^d, \forall a \in C.$$

In particular, taking $a = p_C(x) + d$, with $d \in C_\infty$ yields to $(x - p_C(x))\cdot d \leqslant 0$ for all $d \in C_\infty$ and so $x - p_C(x) \in C_\infty^\circ = \mathcal{Z}_{\bar{\varphi}} \subset \mathcal{Z}$.

Appendix A. Proofs of some technical results

A.1. **Proof of Proposition 2.1.** The proof of Proposition 2.1 is adapted from [3] (paragraph 2.6). First let us see how the recession function c'_∞ can be expressed when c satisfies Assumption (A).

Lemma A.1. *Under Assumption* (A) *it holds*

$$c'_\infty(x,m) = \sup_{k\in\mathbb{N}}\int b_k(x,y)\,m(dy), \qquad x \in \mathcal{X}, m \in \mathcal{M}(\mathcal{Y}).$$

Proof. Since $c(x,\cdot)$ is convex, the function $\lambda \mapsto \frac{c(x,\lambda m)-c(x,0)}{\lambda}$ is non-decreasing on $[0,\infty)$. Therefore, for all $x \in \mathcal{X}$ and $m \in \mathcal{M}(\mathcal{Y})$, it holds

$$c'_\infty(x,m) = \lim_{\lambda\to\infty} \frac{c(x,\lambda m)-c(x,0)}{\lambda} = \sup_{\lambda>0} \frac{c(x,\lambda m)-c(x,0)}{\lambda}. \tag{A.1}$$

Thus, using Assumption (A), it holds

$$\begin{aligned} c'_\infty(x,m) &= \sup_{\lambda>0}\sup_{k\geqslant 0} \frac{\lambda \int b_k(x,y)\,m(dy) + a_k(x) - c(x,0)}{\lambda} \\ &= \sup_{k\geqslant 0}\left\{\int b_k(x,y)\,m(dy) + \sup_{\lambda>0}\frac{a_k(x)-c(x,0)}{\lambda}\right\} \\ &= \sup_{k\geqslant 0}\int b_k(x,y)\,m(dy), \end{aligned}$$

where the last equality comes from the fact that $c(x,0) = \sup_{k\geqslant 0} a_k(x)$, and so for a fixed k, $a_k(x) - c(x,0) \leqslant 0$ and so the function $\lambda \mapsto \frac{a_k(x)-c(x,0)}{\lambda}$ is non-decreasing on $[0,\infty)$. □

We will also need the following lemma

Lemma A.2. *Let λ be a finite measure on $\mathcal{X}$. If $\psi_0,\psi_1,\ldots,\psi_n : \mathcal{X} \to \mathbb{R}$ are λ integrable functions with $\psi_0 \geqslant 0$, then*

$$\int \sup_{0\leqslant k\leqslant n} \psi_k(x)\,\lambda(dx) = \sup_{(f_0,\ldots,f_n)\in\mathcal{F}_n} \int \sum_{k=1}^n \psi_k(x) f_k(x)\,\lambda(dx)$$

where $\mathcal{F}_n$ denotes the set of $n+1$-uples $(f_0,\ldots,f_n)$ of continuous functions $f_0,\ldots,f_n : \mathcal{X} \to [0,1]$ such that $f_0(x)+\cdots+f_n(x) \leqslant 1$.

Proof. See the proof of Proposition 9.4 of [23]. □

Proof of Proposition 2.1. Without loss of generality, one can assume that the functions b_0 and a_0 involved in (A) are nonnegative. If this is not the case, consider the cost function

$$\tilde{c}(x,m) = \sup_{k\geqslant 0}\left\{\int \tilde{b}_k(x,y)\,m(dy) + \tilde{a}_k(x)\right\}, \qquad x\in\mathcal{X}, m\in\mathcal{M}(\mathcal{Y}),$$

with $\tilde{b}_k = b_k - r_0$ and $\tilde{a}_k = a_k - s_0$, with $r_0 = \min_{x\in\mathcal{X},y\in\mathcal{Y}} b_0(x,y)$ and $s_0 = \min_{x\in\mathcal{X}} a_0(x)$. It holds $\tilde{a}_0 \geqslant 0$, $\tilde{b}_0 \geqslant 0$ and one sees that $\tilde{c}(x,m) = c(x,m) - r_0 m(\mathcal{Y}) - s_0$, $x\in\mathcal{X}$, $m\in\mathcal{M}(\mathcal{Y})$, and so $\bar{I}^\mu_{\tilde{c}}[\pi] = \bar{I}^\mu_c[\pi] - (s_0+r_0)$, and $(\mu,\pi)\mapsto \bar{I}^\mu_c[\pi]$ is lower semicontinuous if and only if $(\mu,\pi)\mapsto\bar{I}^\mu_{\tilde{c}}[\pi]$ is.

For $n\geqslant 0$, define for all $\mu\in\mathcal{P}(\mathcal{X})$ and $\pi\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$

$$J^\mu_n[\pi] = \sup_{(f_0,\ldots,f_n)\in\mathcal{F}_n}\left\{\sum_{k=0}^n \iint b_k(x,y) f_k(x)\,\pi(dxdy) + \sum_{k=0}^n \int a_k(x) f_k(x)\,\mu(dx)\right\},$$

where $\mathcal{F}_n$ is defined in Lemma A.2 above. Then consider the functional J^μ defined by

$$J^\mu[\pi] = \sup_{n\geqslant 0} J^\mu_n[\pi], \qquad \pi\in\mathcal{P}(\mathcal{X}\times\mathcal{Y}). \tag{A.2}$$

For each $n\geqslant 0$, the functional $(\mu,\pi)\mapsto J^\mu_n[\pi]$ is lower semicontinuous as a supremum of continuous functionals. Similarly, the functional $(\mu,\pi)\mapsto J^\mu[\pi]$ being the supremum of lower semicontinuous functionals is itself lower semicontinuous.

For all $n\geqslant 0$, write $c_n(x,m) = \sup_{0\leqslant k\leqslant n}\left\{\int b_k(x,y)\,m(dy) + a_k(x)\right\}$, $x\in\mathcal{X}$, $m\in\mathcal{M}(\mathcal{Y})$. According to Lemma A.1, it holds $c'_{n,\infty}(x,m) = \sup_{0\leqslant k\leqslant n}\left\{\int b_k(x,y)\,m(dy)\right\}$, $x\in\mathcal{X}$, $m\in\mathcal{M}(\mathcal{Y})$.

By monotone convergence, it holds

$$\bar{I}_c^\mu[\pi] = \sup_{n \geqslant 0} \bar{I}_{c_n}^\mu[\pi], \qquad \forall \mu \in \mathcal{P}(\mathcal{X}), \forall \pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}).$$

Let us show that for any $\mu \in \mathcal{P}(\mathcal{X})$ and $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, it holds $J_n^\mu[\pi] = \bar{I}_{c_n}^\mu[\pi]$ for all $n \geqslant 0$. This will immediately imply that $\bar{I}_c^\mu[\pi] = J^\mu[\pi]$ for all $\mu \in \mathcal{P}(\mathcal{X})$ and $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ and show the lower semicontinuity of $\bar{I}_c^{\cdot}[\,\cdot\,]$.

Fix $\mu \in \mathcal{P}(\mathcal{X})$; for all $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$, it holds

$$\bar{I}_{c_n}^\mu[\pi] = \int \sup_{0 \leqslant k \leqslant n} \psi_k(x)\,\mu(dx) + \int \sup_{0 \leqslant k \leqslant n} \varphi_k(x)\,\pi_1^s(dx)$$

with, for all $0 \leqslant k \leqslant n$,

$$\psi_k(x) = \frac{d\pi_1^{ac}}{d\mu}(x) \int b_k(x,y)\,\pi^x(dy) + a_k(x), \qquad x \in \mathcal{X}$$

and

$$\varphi_k(x) = \int b_k(x,y)\,\pi^x(dy), \qquad x \in \mathcal{X}.$$

Let $A \subset \mathcal{X}$ be a Borel subset such that $\mu(A) = 0$ and $\pi_1^s(\mathcal{X} \backslash A) = 0$. Define

$$F_k(x) = \begin{cases} \psi_k(x) & \text{if } x \in \mathcal{X}\backslash A \\ \varphi_k(x) & \text{if } x \in A \end{cases}.$$

Then

$$\bar{I}_{c_n}^\mu[\pi] = \int \sup_{0 \leqslant k \leqslant n} F_k(x)\,(\mu(dx) + \pi_1^s(dx)).$$

According to Lemma A.2 above (with $\lambda = \mu + \pi_1^s$), it holds

$$\bar{I}_{c_n}^\mu[\pi] = \int \sup_{0 \leqslant k \leqslant n} F_k(x)\,(\mu(dx) + \pi_1^s(dx)) = \sup_{(f_0,\ldots,f_n) \in \mathcal{F}_n} \int \sum_{k=1}^n F_k(x) f_k(x)\,(\mu(dx) + \pi_1^s(dx)) = J_n^\mu[\pi],$$

which completes the proof. □

A.2. **Proof of Lemma 2.4.** The proof of Lemma 2.4 below is inspired by the proof of [28, Theorem C.12] dealing with entropy type functionals on the space of probability measures.

Lemma A.3. *If $c : \mathcal{X} \times \mathcal{M}(\mathcal{Y}) \to \mathbb{R}$ is convex with respect to its second variable, then for any $x \in \mathcal{X}$ and $m_1, m_2 \in \mathcal{M}(\mathcal{Y})$, it holds*

$$c(x, m_1 + m_2) \leqslant c(x, m_1) + c'_\infty(x, m_2).$$

Proof. Let $\theta \in (0,1)$; using the convexity of $c(x,\cdot)$ and (A.1), one gets

$$\begin{aligned} c(x, m_1 + m_2) &\leqslant \theta c\left(x, \frac{m_1}{\theta}\right) + (1-\theta) c\left(x, \frac{m_2}{1-\theta}\right) \\ &= \theta c\left(x, \frac{m_1}{\theta}\right) + (1-\theta)\left[c\left(x, \frac{m_2}{1-\theta}\right) - c(x,0)\right] + (1-\theta) c(x,0) \\ &\leqslant \theta c\left(x, \frac{m_1}{\theta}\right) + c'_\infty(x, m_2) + (1-\theta) c(x,0), \end{aligned}$$

and the result follows by letting $\theta \to 1$. □

Lemma A.4. *If $c : \mathcal{X} \times \mathcal{M}(\mathcal{Y}) \to \mathbb{R}$ is convex with respect to its second variable and satisfies Assumption (C), then there exists $a, b \geqslant 0$ such that*

$$c(x,m) \leqslant b + a m(\mathcal{Y}), \qquad \forall x \in \mathcal{X}, \forall m \in \mathcal{M}(\mathcal{Y}).$$

Proof. Using (A.1), one gets

$$c(x,m) \leqslant c(x,0) + m(\mathcal{Y})c'_\infty\left(x, \frac{m}{m(\mathcal{Y})}\right) \leqslant c(x,0) + am(\mathcal{Y}),$$

where a is the constant appearing in Assumption (C). Since $c(\,\cdot\,,0)$ is continuous on the compact space $\mathcal{X}$ it is upper bounded by some constant $b \geqslant 0$, which completes the proof. □

Proof of Lemma 2.4. Let $\pi \in \Pi(\mathrm{Supp}(\mu),\nu)$ and denote by $\eta \in \mathcal{P}(\mathrm{Supp}(\mu))$ its first marginal. According to Lemma 2.2, $\pi_n \to \pi$ for the weak topology, and the first marginal of π_n is $\eta_n = K_n\eta$ and is thus absolutely continuous with respect to μ. Since $\pi_n \in \Pi(\ll \mu, \cdot)$, it holds $I_c^\mu[\pi_n] = \bar{I}_c^\mu[\pi_n]$ and since $\bar{I}_c^\mu$ is lower semicontinuous, one gets that

$$\liminf_{n\to\infty} I_c^\mu[\pi_n] = \liminf_{n\to\infty} \bar{I}_c^\mu[\pi_n] \geqslant \bar{I}_c^\mu[\pi].$$

Now let us prove that $\limsup_{n\to\infty} I_c^\mu[\pi_n] \leqslant \bar{I}_c^\mu[\pi]$. Observe that $\pi_n(dxdy) = q_n^x(dy)\mu(dx)$ with

$$q_n^x(dy) = \int K_n(x,z)\pi^z(dy)\eta(dz) = \int K_n(x,z)\pi^z(dy)\eta^{ac}(dz) + \int K_n(x,z)\pi^z(dy)\eta^s(dz) := q_n^{ac,x} + q_n^{s,x}$$

where $\eta = \eta^{ac} + \eta^s$ is the decomposition of η into absolutely continuous and singular parts (with respect to μ). According to Lemma A.3, it holds

$$I_c^\mu[\pi_n] = \int c(x, q_n^x)\,\mu(dx) = \int c(x, q_n^{ac,x} + q_n^{s,x})\,\mu(dx) \leqslant \int c(x, q_n^{ac,x})\,\mu(dx) + \int c'_\infty\left(x, q_n^{s,x}\right)\,\mu(dx).$$

Write $\eta^{ac}(dz) = h(z)\,\mu(dz)$ and let us bound the first term. Since $\int K_n(x,z)\,\mu(dz) = 1$, Jensen inequality yields to

$$\begin{aligned}\int c\left(x, q_n^{ac,x}\right)\,\mu(dx) &= \int c\left(x, \int K_n(x,z)\pi^z(\,\cdot\,)h(z)\mu(dz)\right)\,\mu(dx)\\ &\leqslant \iint K_n(x,z)c\left(x, \pi^z(\,\cdot\,)h(z)\right)\,\mu(dx)\mu(dz)\\ &= \int (K_nC_z)(z)\mu(dz),\end{aligned}$$

where

$$C_z(x) = c\left(x, \pi^z(\,\cdot\,)h(z)\right), \qquad x,z \in \mathcal{X}.$$

By assumption, the function $x \mapsto c\left(x, \pi^z(\,\cdot\,)h(z)\right)$ is continuous on $\mathcal{X}$. Therefore, according to Lemma 2.1, one gets that $K_nC_z(u) \to C_z(u)$ for any $u \in \mathcal{X}$ (and even uniformly in u) as $n \to \infty$. Also, according to Lemma A.4, it holds $c\left(x, \pi^z(\,\cdot\,)h(z)\right) \leqslant b + ah(z)$ and so $K_nC_z(z) \leqslant b + ah(z)$, which is μ integrable. Therefore, according to the dominated convergence theorem, it holds

(A.3)
$$\limsup_{n\to\infty} \int c\left(x, q_n^{ac,x}\right)\,\mu(dx) \leqslant \lim_{n\to\infty} \int (K_nC_z)(z)\mu(dz) = \int C_z(z)\mu(dz) = \int c\left(z, \pi^z(\,\cdot\,)h(z)\right)\,\mu(dz).$$

Now, let us bound the second term. Using the convexity and 1-homogeneity of the function $m \mapsto c'_\infty(x,m)$ one gets

$$\begin{aligned}\int c'_\infty\left(x, q_n^{s,x}\right)\mu(dx) &= \int c'_\infty\left(x, \int K_n(x,z)\pi^z(\,\cdot\,)\eta^s(dz)\right)\mu(dx)\\ &\leqslant \iint c'_\infty\left(x, K_n(x,z)\pi^z(\,\cdot\,)\right)\mu(dx)\eta^s(dz)\\ &= \iint K_n(x,z)c'_\infty\left(x,\pi^z(\,\cdot\,)\right)\mu(dx)\eta^s(dz)\\ &= \int (K_nD_z)(z)\,\eta^s(dz),\end{aligned}$$

where

$$D_z(x) = c'_\infty\left(x,\pi^z(\,\cdot\,)\right), \qquad \forall x,z\in\mathcal{X}.$$

By assumption, the function $x \mapsto c'_\infty\left(x,\pi^z\right)$ is continuous on $\mathcal{X}$. Therefore, according to Lemma 2.1, one gets that $K_nD_z(u)\to D_z(u)$ for any $u\in\mathcal{X}$ (and even uniformly in u) as $n\to\infty$. By assumption, it holds $c'_\infty\left(x,\pi^z\right)\leqslant a$ and so $K_nD_z(z)\leqslant a$. Therefore, according to the dominated convergence theorem, it holds

$$\limsup_{n\to\infty}\int c\left(x,q_n^{s,x}\right)\mu(dx) \leqslant \lim_{n\to\infty}\int (K_nD_z)(z)\,\eta^s(dz) = \int D_z(z)\eta^s(dz) = \int c'_\infty(z,\pi^z)\eta^s(dz). \tag{A.4}$$

Adding (A.3) and (A.4), one gets that $\limsup_{n\to\infty} I_c^\mu[\pi_n]\leqslant \bar{I}_c^\mu[\pi]$, which completes the proof. □

References

1. A. Alfonsi, J. Corbetta, and B. Jourdain, *Sampling of probability measures in the convex order by Wasserstein projection*, Ann. Inst. Henri Poincaré Probab. Stat. **56** (2020), no. 3, 1706–1729. MR 4116705 2
2. J.-J. Alibert, G. Bouchitté, and T. Champion, *A new class of costs for optimal transport planning*, Eur. J. Appl. Math. **30** (2019), no. 6, 1229–1263 (English). 2, 20
3. L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 2000. 15, 43
4. J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer, *Existence, duality, and cyclical monotonicity for weak transport costs*, Calc. Var. Partial Differ. Equ. **58** (2019), no. 6, 28 (English), Id/No 203. 2, 3, 20, 37
5. J. Backhoff-Veraguas, M. Beiglböck, and G. Pammer, *Weak monotone rearrangement on the line*, Electron. Commun. Probab. **25** (2020), Paper No. 18, 16. MR 4069738 2
6. J. Backhoff-Veraguas and G. Pammer, *Applications of weak transport theory*, Bernoulli **28** (2022), no. 1, 370–394. MR 4337709 2
7. ———, *Stability of martingale optimal transport and weak optimal transport*, Ann. Appl. Probab. **32** (2022), no. 1, 721–752. MR 4386541 20
8. M. Beiglböck, P. Henry-Labordère, and F. Penkner, *Model-independent bounds for option prices—a mass transport approach*, Finance Stoch. **17** (2013), no. 3, 477–501. MR 3066985 2
9. M. Beiglböck and N. Juillet, *On a problem of optimal transport under marginal martingale constraints*, Ann. Probab. **44** (2016), no. 1, 42–106. MR 3456332 2
10. J.-D. Benamou, G. Carlier, M. Cuturi, L. Nenna, and G. Peyré, *Iterative Bregman projections for regularized transportation problems*, SIAM J. Sci. Comput. **37** (2015), no. 2, A1111–A1138. MR 3340204 2
11. Y. Brenier, *Décomposition polaire et réarrangement monotone des champs de vecteurs*, C. R. Acad. Sci. Paris Sér. I Math. **305** (1987), no. 19, 805–808. MR 923203 2, 25, 42
12. ———, *Polar factorization and monotone rearrangement of vector-valued functions*, Comm. Pure Appl. Math. **44** (1991), no. 4, 375–417. MR 1100809 2, 42
13. L. A. Caffarelli, *Monotonicity properties of optimal transportation and the FKG and related inequalities*, Comm. Math. Phys. **214** (2000), no. 3, 547–563. MR 1800860 2
14. E. Cazelles, F. Tobar, and J. Fontbona, *A novel notion of barycenter for probability distributions based on optimal weak mass transport*, 2022. 2
15. P. Choné and F. Kramarz, *Matching Workers' Skills and Firms' Technologies: From Bundling to Unbundling*, Preprint, 2021. 1, 3, 4, 6, 12, 35

16. M. Cuturi, *Sinkhorn distances: Lightspeed computation of optimal transport.*, in Advances in Neural Information Processing Systems (NIPS) (2013), no. 26, 2292–2300. 2
17. M. Fathi, N. Gozlan, and M. Prod'homme, *A proof of the Caffarelli contraction theorem via entropic regularization*, Calc. Var. Partial Differential Equations **59** (2020), no. 3, Paper No. 96, 18. MR 4098037 2
18. A. Galichon, *Optimal transport methods in economics*, Princeton University Press, Princeton, NJ, 2016. MR 3586373 1
19. A. Galichon, P. Henry-Labordère, and N. Touzi, *A stochastic control approach to no-arbitrage bounds given marginals, with an application to lookback options*, Ann. Appl. Probab. **24** (2014), no. 1, 312–336. MR 3161649 2
20. N. Gozlan and N. Juillet, *On a mixture of Brenier and Strassen theorems*, Proc. Lond. Math. Soc. (3) **120** (2020), no. 3, 434–463 (English). 2, 5, 37, 42, 43
21. N. Gozlan and C. Léonard, *Transport inequalities. A survey*, Markov Process. Related Fields **16** (2010), no. 4, 635–736. MR 2895086 2
22. 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. Henri Poincaré Probab. Stat. **54** (2018), no. 3, 1667–1693. MR 3825894 2
23. N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali, *Kantorovich duality for general transport costs and applications*, J. Funct. Anal. **273** (2017), no. 11, 3327–3405 (English). 1, 2, 44
24. J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of convex analysis*, Grundlehren Text Editions, Springer-Verlag, Berlin, 2001. 41, 43
25. Y.-H. Kim and Y. L. Ruan, *Backward and forward wasserstein projections in stochastic order*, 2021. 5
26. M. Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. MR 1849347 2
27. C. Léonard, *A survey of the Schrödinger problem and some of its connections with optimal transport*, Discrete Contin. Dyn. Syst. **34** (2014), no. 4, 1533–1574. MR 3121631 2
28. J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. Math. (2) **169** (2009), no. 3, 903–991 (English). 14, 45
29. K. Marton, *Bounding $\bar{d}$-distance by informational divergence: a method to prove measure concentration*, Ann. Probab. **24** (1996), no. 2, 857–866. 2
30. ———, *A measure concentration inequality for contracting Markov chains*, Geom. Funct. Anal. **6** (1996), no. 3, 556–571. 2
31. G. Peyré and M. Cuturi, *Computational optimal transport: With applications to data science*, vol. 11, Now Publishers, Inc., 2019. 1
32. R. T. Rockafellar, *Convex analysis*, Princeton Mathematical Series, No. 28, Princeton University Press, Princeton, N.J., 1970. MR 0274683 29
33. V. Strassen, *The existence of probability measures with given marginals*, Ann. Math. Statist. **36** (1965), 423–439. 4, 30
34. C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003. 1
35. ———, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009, Old and new. 1
36. C. Zălinescu, *Convex analysis in general vector spaces*, Singapore: World Scientific, 2002 (English). 19

Philippe Choné : CREST-ENSAE, Institut Polytechnique de Paris, France

Email address: Philippe.Chone@ensae.fr

Nathael Gozlan : Université Paris Cité, MAP5, UMR 8145, 45 rue des Saints Pères, 75270 Paris Cedex 06

Email address: natael.gozlan@parisdescartes.fr

Francis Kramarz : CREST-ENSAE, Institut Polytechnique de Paris, France and Department of Economics, Uppsala University, Sweden

Email address: Francis.Kramarz@ensae.fr