HAL open science

ON A MIXTURE OF BRENIER AND STRASSEN THEOREMS

Nathael Gozlan, Nicolas Juillet

To cite this version:

Nathael Gozlan, Nicolas Juillet. ON A MIXTURE OF BRENIER AND STRASSEN THEOREMS. Proceedings of the London Mathematical Society, 2020, 10.1112/plms.12302 . hal-01855463v2

HAL Id: hal-01855463

https://hal.science/hal-01855463v2

Submitted on 21 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ON A MIXTURE OF BRENIER AND STRASSEN THEOREMS

NATHAEL GOZLAN AND NICOLAS JUILLET

Abstract. We give a characterization of optimal transport plans for a variant of the usual quadratic transport cost introduced in [33]. Optimal plans are composition of a deterministic transport given by the gradient of a continuously differentiable convex function followed by a martingale coupling. We also establish some connections with Caffarelli's contraction theorem [14].

1. Introduction

Given two probability measures μ, ν on $\mathbb{R}^d$, we recall that a *transport plan* between μ and ν is a probability measure π on $\mathbb{R}^d \times \mathbb{R}^d$ such that

$$\pi(A \times \mathbb{R}^d) = \mu(A) \qquad \text{and} \qquad \pi(\mathbb{R}^d \times B) = \nu(B),$$

for all Borel sets A, B of $\mathbb{R}^d$. The set of all transport plans between μ and ν will be denoted by $C(\mu, \nu)$ in all the paper. It will be convenient to represent a transport plan $\pi \in C(\mu, \nu)$ in the following disintegrated form

$$d\pi(x, y) = d\mu(x) dp_x(y), \tag{1}$$

where $x \mapsto p_x$ is a (μ-almost surely unique) probability kernel called a *transport kernel*. A couple (X, Y) of random variables such that $X \sim \mu$ and $Y \sim \nu$ is called a *coupling* of μ and ν. By a slight abuse of vocabulary, we also call couplings of μ and ν the transport plans of $C(\mu, \nu)$, which also explains the notation with letter C. In all the paper, $|\cdot|$ will denote the standard Euclidean norm on $\mathbb{R}^d$. The corresponding scalar product will be denoted by $x \cdot y$ or $\langle x, y\rangle$ when the expression of x or y is too long. For $k \geq 1$, we will denote by $\mathcal{P}_k(\mathbb{R}^d)$ the *Wasserstein space* of order k, i.e the set of probability measures μ such that $\int |x|^k \, d\mu(x) < \infty$.

1.1. Existence of structured couplings between probability measures.

The question to construct couplings between probability measures having some nice structures or properties is natural both in probability theory and analysis. Let us recall two famous results guarantying the existence of couplings with rigid but very different structures, namely the theorems of Brenier and Strassen.

A fundamental result by Brenier [12, 13] (see [55, §1.3] for a discussion of the bibliography around this result) shows that if μ is say absolutely continuous with respect to Lebesgue measure, then there exists some convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$ such that $\nabla\varphi$ pushes forward μ onto ν. In other words, there exists a *deterministic* coupling π° of the form

$$d\pi^\circ(x, y) = d\mu(x)\delta_{\nabla\varphi(x)}(y).$$

Moreover, assuming in addition that μ and ν belong to $\mathcal{P}_2(\mathbb{R}^d)$, this π° is the unique optimal coupling in the Monge-Kantorovich transport problem for the quadratic cost:

$$\iint |x - y|^2 \, d\pi^\circ(x, y) = \inf_{\pi \in C(\mu,\nu)} \iint |x - y|^2 \, d\pi(x, y).$$

Date: December 21, 2018.
1991 *Mathematics Subject Classification.* 60A10, 49J55, 60G42.
Key words and phrases. Optimal transport, Martingales, Brenier Theorem, Strassen Theorem.

This result has then been extended to various cost functions [26] and state spaces [54, 28, 4, 24, 23, 21, 9, 29] and has had numerous applications in PDE, geometric analysis or probability theory, see [3, 6, 72, 73] and the references therein.

On the other hand, from a probabilistic point of view it is natural to investigate the existence of *martingale couplings* between elements of $\mathcal{P}_1(\mathbb{R}^d)$, that is to say to look for couplings $\pi \in C(\mu,\nu)$ which correspond to the law of a martingale $(X_t)_{t\in\{0;1\}}$. This martingale requirement means that the kernel p appearing in (1) satisfies

$$\int y\, dp_x(y) = x, \quad \text{for } \mu \text{ almost every } x \in \mathbb{R}^d.$$

As for the Brenier Theorem, it turns out that the existence of a martingale coupling between μ and ν is not automatic. Given $\mu,\nu \in \mathcal{P}_1(\mathbb{R}^d)$, a necessary and sufficient condition is given by Strassen Theorem [69]: there exists a martingale coupling between μ and ν if and only if μ is dominated by ν in the convex order. Recall that if $\mu,\nu \in \mathcal{P}_1(\mathbb{R}^d)$, one says that μ is dominated by ν in the convex order if $\int f\, d\mu \le \int f\, d\nu$ for all convex functions $f : \mathbb{R}^d \to \mathbb{R}$. [1] This is denoted as follows in the what follows: $\mu \preceq_c \nu$. Note that Strassen's result has been generalized at the time continuous process level by Kellerer [46, 47] (see also [8]): if a family $(\mu_t)_{t\ge 0}$ of elements of $\mathcal{P}_1(\mathbb{R})$ is increasing for the convex order (a so-called *peacocks* using the terminology of [40]), then there exists a process $(X_t)_{t\ge 0}$ which is together *Markovian* and a martingale such that $X_t \sim \mu_t$ for all $t \ge 0$. See the monograph [37] on peacocks, [51, 39] for the approach by Lowther and [44, 38, 10] for extensions.

1.2. **In between Brenier and Strassen.** Given two probability measures μ,ν on $\mathbb{R}^d$ having their first moments finite, it is not always possible to go from μ to ν using a deterministic mapping given by the gradient of a convex function, nor to go from μ to ν using a martingale coupling. In this paper we will be interested in couplings obtained by composition of these two classical transport methods, which as we will see in the next paragraph always exist.

More precisely, we will look for couplings (X,Y) of μ and ν such that there exists a convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$ of class $\mathcal{C}^1$ such that

$$\mathbb{E}[Y|X] = \nabla\varphi(X), \qquad \text{a.s.} \tag{2}$$

Introducing $X' = \nabla\varphi(X)$, we can see (X,Y) as the initial and final states of a (time inhomogeneous) Markov chain (X,X',Y) where the first transition step is deterministic and given by the gradient of a convex function and the second transition step is a martingale. As mentioned above, such couplings always exist: simply choosing X and Y independent gives a trivial example corresponding to $\varphi(x) = x \cdot \mathbb{E}[Y]$, $x \in \mathbb{R}^d$.

The purpose of our main result (Theorem 1.2 below) is therefore to distinguish some *special* couplings (X,Y) among those satisfying (2) by showing that they are solutions of an optimal transport problem that we shall now present.

Let $\theta : \mathbb{R}^d \to \mathbb{R}^+$ be some convex function and $\mu,\nu \in \mathcal{P}_1(\mathbb{R}^d)$; using the terminology of [33], the barycentric optimal transport cost between μ and ν is defined by

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \inf_{\pi\in C(\mu,\nu)} \int \theta\left(\int y\, dp_x(y) - x\right) d\mu(x), \tag{3}$$

where p is the kernel appearing in the disintegration formula (1) for π. Notice that the bar over $\mathcal{T}_\theta$ in (3) stands for the beginning of the word *bar*ycentric. In probabilistic notations, this optimal transport cost can be expressed as follows:

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \inf \mathbb{E}\left[\theta\left(\mathbb{E}[Y|X] - X\right)\right]$$

[1] Note that since a convex function f on $\mathbb{R}^d$ is always bounded from below by some affine function, the integral of f with respect to a probability measure with finite first moment always makes sense in $\mathbb{R} \cup \{+\infty\}$.

where the infimum runs over the set of random vectors (X, Y) such that $X \sim \mu$ and $Y \sim \nu$. If $d\pi(x, y) = d\mu(x)dp_x(y) \in C(\mu, \nu)$ is such that $\overline{\mathcal{T}}_\theta(\nu|\mu) = \int \theta\left(\int y\, dp_x(y) - x\right) d\mu(x)$, we call it an *optimal* transport plan (or coupling) from μ to ν (for the cost $\overline{\mathcal{T}}_\theta$). These barycentric costs are actually part of a more general family of transport costs, called "weak transport costs" or "generalized transport costs", introduced under the first name by the first author together with Roberto, Samson and Tetali in [33] (see Section 1.4 below for a more complete presentation).

In this paper, we will mainly focus on the barycentric transport cost, denoted by $\overline{\mathcal{T}}_2$ in the remainder of the present text, in relation to $\overline{\mathcal{T}}_\theta$ defined in (3), as it is associated to the function $\theta(u) = |u|^2$, $u \in \mathbb{R}^d$. For all $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$

$$\overline{\mathcal{T}}_2(\nu|\mu) = \inf_{\pi \in C(\mu,\nu)} \int \left| \int y\, dp_x(y) - x \right|^2 d\mu(x).$$

In [33, Theorem 2.11], the following Kantorovich type duality formula has been obtained:

Theorem 1.1. *If $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^d)$ then*

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \sup_f \left\{ \int Q_\theta f\, d\mu - \int f\, d\nu \right\},$$

where

$$Q_\theta f(x) = \inf_{y \in \mathbb{R}^d} \{f(y) + \theta(y - x)\}, \qquad x \in \mathbb{R}^d,$$

and where the supremum runs over the set of all functions f which are convex*, Lipschitz and bounded from below.*

Note that $Q_\theta(f)$ is the c-transform of f related to the cost $c(x, y) = \theta(y-x)$, $x, y \in \mathbb{R}^d$, as defined in optimal transport theory (see for instance [73, Definition 5.2]). The difference with respect to the usual duality formula for the global transport cost associated to θ (see [73, Theorem 5.10]) is that we optimize over a class of convex functions. Several applications of this duality formula were already investigated in [33], mainly in connection with transport-entropy inequalities and deviation inequalities for convex functions. In particular, let us mention that Strassen Theorem can be derived from this duality theorem, see [33, Section 3]. Here we shall use this duality result to describe optimal transport plans for $\overline{\mathcal{T}}_2$.

Before stating our main result, we need some preparation. First recall that the transport distance W_2 is defined for all probability measures $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ by

$$W_2^2(\mu, \nu) = \inf_{\pi \in C(\mu,\nu)} \iint |y - x|^2\, d\pi(x, y).$$

Given a probability $\nu \in \mathcal{P}_1(\mathbb{R}^d)$, we denote by

$$B_\nu = \{\eta \in \mathcal{P}_1(\mathbb{R}^d) : \eta \preceq_c \nu\}$$

the set of all probability measures which are dominated by ν in the convex order. This set B_ν is easily seen to be convex in the usual sense and the proof of Proposition 1.1 will show that (generalized) W_2-geodesics with endpoints in B_ν are contained in B_ν.

Proposition 1.1. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$. There exists a unique probability measure $\bar{\mu} \in B_\nu$ such that*

$$W_2(\bar{\mu}, \mu) = \inf_{\eta \in B_\nu} W_2(\eta, \mu) = \overline{\mathcal{T}_2}(\nu|\mu).$$

We call $\bar{\mu}$ the projection of μ on B_ν.

Remark 1.1. *The identity*

$$\overline{\mathcal{T}}_2(\nu|\mu) = \inf_{\eta \in B_\nu} W_2^2(\eta, \mu)$$

of Proposition 1.1 was already observed in [31, Proposition 3.1] *in dimension* 1 *when* μ has no atom *and then generalized to higher dimensions in* [65, Proposition 4.1] *when* μ *is* absolutely continuous *with respect to Lebesgue. After a first version of this work has been released, we learned that Proposition 1.1 has been independently obtained by Alfonsi, Corbetta and Jourdain in* [1] *in connections with the study of algorithms approximating the martingale transport problem.*

We recall that if $g : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is some convex function, the Legendre transform $g \mapsto g^*$ is defined by

$$g^*(y) = \sup_{x \in \mathbb{R}^d} \{x \cdot y - g(x)\}, \qquad y \in \mathbb{R}^d.$$

We will keep on using notation "$*$" for the Legendre transform all along the paper. With these notions in hand, we can now state the main result of this paper which describes the set of all optimal couplings for $\overline{\mathcal{T}}_2$.

Theorem 1.2. *Let* $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$.

(a) There exists some lower semi-continuous convex function $f^\circ : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$, *which is integrable with respect to* ν *and such that*

$$\overline{\mathcal{T}_2}(\nu|\mu) = \int Q_2 f^\circ \, d\mu - \int f^\circ \, d\nu. \tag{4}$$

In this equation Q_2 *corresponds to* Q_θ, *defined in Theorem 1.1, where* θ *is the quadratic cost. It is given for any function* $g : \mathbb{R}^d \to \mathbb{R}$ *by*

$$Q_2 g(x) = \inf_{y \in \mathbb{R}^d} \{g(y) + |y - x|^2\}, \qquad x \in \mathbb{R}^d.$$

(b) Let h *and* φ *be the convex functions defined by*

$$h(x) = \frac{f^\circ(x) + |x|^2}{2}, \quad x \in \mathbb{R}^d \qquad \text{and} \qquad \varphi(y) = h^*(y), \quad y \in \mathbb{R}^d.$$

The function φ *is* $\mathcal{C}^1$*-smooth on* $\mathbb{R}^d$ *and the map* $\nabla\varphi$ *is* 1*-Lipschitz on* $\mathbb{R}^d$. *The projection* $\bar{\mu}$ *of* μ *on* B_ν *is such that*

$$\bar{\mu} = \nabla\varphi_{\#}\mu.$$

(c) The set of optimal transport plans for $\overline{\mathcal{T}}_2(\nu|\mu)$ *is non-empty. Moreover if* (X, Y) *is a coupling of* μ, ν *such that* $\overline{\mathcal{T}}_2(\nu|\mu) = \mathbb{E}[|\mathbb{E}[Y|X] - X|^2]$, *then* $\mathbb{E}[Y|X]$ *has law* $\bar{\mu}$, $\mathbb{E}[Y|X] = \nabla\varphi(X)$ *almost surely and* $(\mathbb{E}[Y|X], Y)$ *is a martingale (this last point being of course always true).*

Remark 1.2.

- *The fact that* φ *is convex and* $\bar{\mu} = \nabla\varphi_{\#}\mu$ *implies that* $x \mapsto \nabla\varphi(x)$ *provides the optimal transport plan* $\pi = (\mathrm{Id} \times \nabla\varphi)_{\#}\mu \in C(\mu, \bar{\mu})$ *for the quadratic cost. In other words, if* X *has law* μ *then*

$$W_2^2(\bar{\mu}, \mu) = \mathbb{E}[|\nabla\varphi(X) - X|^2]. \tag{5}$$

  As in the classical optimal transport theory the conjugate potential $\psi = \varphi^*(= h)$ *is such that* $x \in \partial\psi(y)$, π*-almost surely. Here* $\partial\psi$ *is the subgradient of* ψ.
- *We underline the fact that there is no assumption on the measure* μ *(as for instance the condition that it does not give mass to small sets, which is classical for Brenier transport, see* [28] *for a minimal condition). As the potential function* φ *is* $\mathcal{C}^1$*-smooth its gradient is well defined everywhere. Hence, the optimal transport map* $x \mapsto \nabla\varphi(x)$ *is pointwise well-defined and there is no ambiguity on the range of small sets.*
- *We stress that the optimal transport map from* μ *to* $\bar{\mu}$ *is continuous and even* 1*-Lipschitz continuous, which is not automatic for quadratic optimal transport between arbitrary measures of* $\mathcal{P}_2(\mathbb{R}^d)$.

- *In a previous version of this paper, Theorem 1.2 appeared with the additional assumption that μ and ν are compactly supported. Soon after this first version has been released, a paper by Backhoff-Veraguas, Beiglböck and Pammer [5] proposed an improved version of Items (b) and (c) of our main result* removing this compactness assumption. *Their approach is based on a clever combination of generalized cyclical monotonicity arguments and on the first compact version of our Theorem 1.2. In the present version of our paper, we improve our preceding proof of Item (a) of Theorem 1.2 in order to remove also the compactness of supports assumption (the proof of Items (b) and (c) being essentially unchanged). The main difference between the previous proof and the present one is that the existence of the dual optimizer is obtained using some version of Ascoli Theorem adapted to convex functions instead of the usual version. To keep visible the incremental progresses around Theorem 1.2, we have treated separately the compact and the general cases in the proof Section 6.*

Before presenting the organization of the paper, let us mention that the methods and results developed in this work are likely to extend to more general barycentric or even some classes of weak optimal transport problems, see §1.4 for a presentation of the weak transport costs. We focus on the particular case of $\overline{\mathcal{T}}_2$ to avoid unnecessary generality and make the reading of this paper more fluid.

1.3. **Other results and organization of the paper.** Let us now describe the content of the paper.

Section 2 adresses the question of characterizing equality cases between W_2^2 and $\overline{\mathcal{T}}_2$. The second main result of the paper, Theorem 2.1, states that $W_2^2(\nu,\mu) = \overline{\mathcal{T}}_2(\nu|\mu)$ *if and only if* there exists some continuously differentiable convex function φ such that $\nabla\varphi$ is 1-Lipschitz and $\nu = \nabla\varphi_{\#}\mu$. In other words, this result shows that the Brenier map from μ to ν is a contraction, i.e a 1-Lipschitz map, if and only if $\bar{\mu} = \nu$. This observation is then compared to a classical theorem by Caffarelli [14] that states that probability measures with a log-concave density with respect to the standard Gaussian measure are contractions of it.

In Section 3, we generalize the notion of $c-$monotonicity to the case of the transport cost $\overline{\mathcal{T}}_2$. We illustrate the power of this notion, by giving a new proof of Strassen Theorem for submartingales in dimension 1.

In Section 4, we study a concrete example and describe the function $\nabla\varphi$ and the projected measure $\bar{\mu}$ when μ is some arbitrary probability measure on $\mathbb{R}^d$ (with finite second moment) and ν is a discrete probability measure concentrated on the vertices of a simplex in $\mathbb{R}^d$.

Section 5 recalls some other examples and results appearing in the recent literature.

Finally, Section 6 contains the postponed proofs of Proposition 1.1 and Theorem 1.2 (for both the compact and general cases) and of some auxiliary lemmas.

1.4. **More about weak optimal transport costs.** As mentioned at the end of §1.2, the barycentric transport costs (3) enter a more general family of transport costs, that we will now recall.

Let $(\mathcal{X}, d)$ be a Polish space and $\mathcal{P}(\mathcal{X})$ be the set of Borel probability measures on $\mathcal{X}$. Given a cost function $c : \mathcal{X} \times \mathcal{P}(\mathcal{X}) \to \mathbb{R}^+$ and probability measures μ, ν on $\mathcal{X}$ the optimal weak transport cost $\mathcal{T}_c(\nu|\mu)$ from μ to ν is defined by

$$\mathcal{T}_c(\nu|\mu) = \inf_{\pi \in C(\mu,\nu)} \int c(x, p_x)\, d\mu(x),$$

where π and p are related together by (1). Note that this definition makes sense under mild regularity assumptions on c, that we will not detail in this exposition.

On the one hand, when $c(x,p) = \int \rho(x,y)\,dp(y)$ for some measurable non-negative function ρ on $\mathcal{X}^2$, one recovers the usual Monge-Kantorovich optimal transport cost

$$\mathcal{T}_\rho(\mu,\nu) = \inf_{\pi \in C(\mu,\nu)} \int \rho(x,y)\,d\pi(x,y).$$

On the other hand, the optimal barycentric transport costs presented in (3) corresponds to cost functions of the form $c(x,p) = \theta(\int y\,dp(y) - x)$.

The first appearance of this form of transport cost goes back to the works of Marton [53, 52] and Talagrand [70, 71] in connections with the concentration of measure phenomenon via the so-called transport-entropy inequalities (see [48] for an introduction to concentration of measure phenomenon and [30] for a survey on transport-entropy inequalities). For instance, Marton considers in [52] the following simple cost function:

$$c_{\mathrm{Marton}}(x,p) = \left(\int \mathbf{1}_{x\neq y}\,dp(y)\right)^2, \qquad x \in \mathcal{X},$$

and recovers a fundamental concentration of measure result by Talagrand [70] for product probability measures. After Marton, some other universal concentration of measures results by Talagrand were recovered or improved by Dembo [20] and Samson [62, 63, 64] using variants of Marton's cost c_{Marton}. Motivated by questions related to concentration and curvature properties of discrete measures, the paper [33] introduces the general definition given above and studies some of the properties of this new class of transport costs. In particular Kantorovich type duality formulas are obtained [33, Theorem 9.6] under the assumption that c is convex with respect to the p variable (and some additional mild regularity conditions). We refer to [32, 31, 66, 68, 22] for works directly connected to [33] and to [65] for an up-to-date survey of applications of weak transport costs to concentration of measure.

Besides their many applications in the field of functional inequalities and concentration of measure, it turns out that weak transport costs are also interesting in themselves as a natural generalization of the transportation problem. Indeed, the main interest of this definition is that it enables the introduction of additional constraints to the transfers of mass. For instance, a cost function of the form

$$c(x,p) = \begin{cases} \int \rho(x,y)\,dp(y) & \text{if } \int y\,dp(y) = x \\ +\infty & \text{otherwise} \end{cases}$$

where $\rho : \mathbb{R}^d \times \mathbb{R}^d$ is some non-negative measurable function gives back the notion of optimal transport with martingale constraint: if $\mu \preceq_c \nu$ (say compactly supported to avoid definition problem), then

$$\mathcal{T}_c(\nu|\mu) = \inf\{\mathbb{E}[\rho(X,Y)] : X \sim \mu, Y \sim \nu, (X,Y) \text{ is a martingale}\}.$$

This optimal transport problem with martingale contraint has been thoroughly studied in [7] for the dimension 1. There, the martingale transport problem is studied for particular families of costs satisfying the cross derivative condition $\partial_x \partial_y^2 \rho < 0$ giving rise to the left-curtain coupling (on this coupling, see also [35, 45, 43]) and for the cost functions $\rho : (x,y) \mapsto \pm|y-x|$, generalizing results by Hobson and his coauthors, Neuberger and Klimmek [42, 41], respectively. The supermartingale problem in dimension 1 with cross-derivative condition is studied in [57]. In higher dimension $\rho : (x,y) \mapsto \pm\|y-x\|$ is studied in [27, 50], more general costs are considered in [19]. We refer to [2] for existence results in the dual optimal transport problem with martingale contraints making use of the formulation in terms of weak costs as above.

As observed by Conforti [15] and Conforti and Ripani [16] the class of weak transport costs also entails the notion of entropic costs related to the Schrödinger problem (see the nice survey by C. Léonard [49]). We also refer to the recent articles by Alibert,

Bouchitté and Champion [2] (see also Section 5) and Bowles and Ghoussoub [11] for further developments and examples.

2. Link with Caffarelli's contraction theorem

In this section we admit the main theorem of this paper, Theorem 1.2, and derive a link with the celebrated Caffarelli's contraction theorem of [14]. Precisely we investigate the question of determining in which case

$$W_2^2(\nu,\mu) = \overline{\mathcal{T}}_2(\nu|\mu)$$

and we show how this question is related (when μ is the standard Gaussian) to Caffarelli's result. The following result is the second main contribution of the paper:

Theorem 2.1. *Let $\mu,\nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mathbb{R}^d$ equipped with the standard Euclidean norm. The following statements are equivalent:*

(a) There exists a convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$ of class $\mathcal{C}^1$ such that $\nabla\varphi$ is 1-Lipschitz such that $\nu = \nabla\varphi_{\#}\mu$.
(b) The projection $\bar{\mu}$ of μ on the set $B_\nu = \{\eta \in \mathcal{P}_1(\mathbb{R}^d) : \eta \preceq_c \nu\}$ is ν.
(c) It holds $W_2^2(\nu,\mu) = \overline{\mathcal{T}}_2(\nu|\mu)$.

In dimension 1, the equivalence between (c) and (a) has been obtained by Shu in a slightly different form in [67].

Let us denote by γ_d the standard Gaussian measure on $\mathbb{R}^d$ and recall the statement of Caffarelli's contraction theorem.

Theorem 2.2 (Caffarelli's contraction theorem [14])**.** *If ν is a probability measure with a density with respect to γ_d of the form e^{-V}, with $V : \mathbb{R}^d \to \mathbb{R}\cup\{\infty\}$ a convex function, then there exists a convex function φ of class $\mathcal{C}^1$ on $\mathbb{R}^d$ such that the Brenier transport map $\nabla\varphi$ from γ_d to ν is 1-Lipschitz.*

Theorem 2.2 is an important result with numerous applications in the field of functional inequalities [34, 18, 56]. Note that the assumption on ν can equivalently be formulated as: ν has density e^{-U} with respect to the Lebesgue measure where $x \mapsto U(x) - |x|^2/2 \in \mathbb{R}\cup\{\infty\}$ is convex. Such measures are called 1-*uniformly log-concave* in the literature.

The following corollary is an immediate consequence of Theorems 2.1 and 2.2.

Corollary 2.1. *If ν is a probability measure with a density with respect to γ_d of the form e^{-V}, with $V : \mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$ a convex function, then the projection $\overline{\gamma_d}$ of γ_d on B_ν is equal to ν.*

Remark 2.1. *Let $d\nu = e^{-V}\, d\gamma_d$ with V convex on $\mathbb{R}^d$.*

- *According to Theorem 2.1 the conclusion of Theorem 2.2 is logically equivalent to the statement $\overline{\gamma_d} = \nu$. It would be very interesting to prove directly that $\overline{\gamma_d} = \nu$ since this would give an alternative proof Theorem 2.2. This question will be considered elsewhere.*
- *To complete the picture, let us mention a nice result of Hargé [34] that states that if in addition $\int x\, d\nu(x) = 0$, then $\nu \preceq_c \gamma_d$.*

Now let us turn to the proof of Theorem 2.1. We will need the following classical lemma (see [36, Theorem E 4.2.1]) whose proof is recalled in Section 6 for the sake of completeness.

Lemma 2.1. *Let $g : \mathbb{R}^d \to \mathbb{R}\cup\{+\infty\}$ be a lower semi-continuous convex function such that $g(x_o) < +\infty$ for some $x_o \in \mathbb{R}^d$ and recall $g^*(y) = \sup_{x\in\mathbb{R}^d}\{x\cdot y - g(x)\}$, $y \in \mathbb{R}^d$. The following are equivalent:*

(a) The function $x \mapsto g(x) - \frac{|x|^2}{2}$ is convex on $\mathbb{R}^d$.
(b) The function $y \mapsto \frac{|y|^2}{2} - g^*(y)$ is convex and finite valued on $\mathbb{R}^d$.

(c) The function g^* is of class $\mathcal{C}^1$ and ∇g^* is 1-Lipschitz on $\mathbb{R}^d$.

Notice that (a) quantitatively states that g is more convex than $x \mapsto |x|^2/2$ while (b) states that g^* is less convex than $x \mapsto |x|^2/2$. Now let us prove Theorem 2.1.

Proof of Theorem 2.1. Let us show that (a) implies (b). Define $h = \varphi^*$ and $f = 2\varphi^* - |x|^2$. According to Lemma 2.1 (the implication $(c) \Rightarrow (a)$ applied with $g = \varphi^*$), we have that f is convex. An easy calculation shows that $Q_2 f(x) = |x|^2 - 2\varphi(x)$, for every $x \in \mathbb{R}^d$, and moreover it holds $\varphi^*(\nabla\varphi(x)) = x \cdot \nabla\varphi(x) - \varphi(x)$, $x \in \mathbb{R}^d$. This together with the fact that $\nabla\varphi$ sends μ onto ν yields that

$$\begin{aligned}
W_2^2(\nu,\mu) &\leq \int |\nabla\varphi(x) - x|^2 \, d\mu(x) \\
&= \int |x|^2 \, d\mu(x) + \int |y|^2 \, d\nu(y) - 2\int x \cdot \nabla\varphi(x) \, d\mu(x) \\
&= \int |x|^2 \, d\mu(x) + \int |y|^2 \, d\nu(y) - 2\int \varphi(x) + \varphi^*(\nabla\varphi(x)) \, d\mu(x) \\
&= \int Q_2 f \, d\mu - \int f \, d\nu \\
&\leq \overline{\mathcal{T}}_2(\nu|\mu) \\
&\leq W_2^2(\nu,\mu),
\end{aligned}$$

where the second inequality comes from the convexity of f and the duality formula for $\overline{\mathcal{T}}_2$, see Theorem 1.2 (a). Note that, according to Lemma 2.1 (the implication $(c) \Rightarrow (b)$ with $g = \varphi^*$) we have that $y \mapsto |y|^2/2 - \varphi(y)$ is convex. Therefore it is bounded from below by some affine function and so φ is bounded from above by some quadratic function, which shows that φ is integrable with respect to μ. Since $x \mapsto \varphi(x) + \varphi^*(\nabla\varphi(x))$ is integrable with respect to μ, one concludes that $\varphi^*(\nabla\varphi(x))$ is also integrable with respect to μ, which enables to split the integral at the fourth line. This also shows that f is integrable with respect to ν.

Therefore, the function f is optimal for the dual problem and so, according to Theorem 1.2, $\nu = \nabla h^*_{\#}\mu = \nabla\varphi_{\#}\mu = \bar{\mu}$, which shows (b). Now according to Theorem 1.2, $W_2^2(\bar{\mu},\mu) = \overline{\mathcal{T}}_2(\nu|\mu)$ so (b) implies (c). On the other hand, if $W_2^2(\nu,\mu) = \overline{\mathcal{T}}_2(\nu|\mu)$, then according to Proposition 1.1 and Theorem 1.2, it holds $\bar{\mu} = \mu$, so (c) implies (b) as well. Finally, according to Theorem 1.2, there always exist a convex function φ of class $\mathcal{C}^1$ on $\mathbb{R}^d$ such that $\nabla\varphi$ is 1-Lipschitz and $\bar{\mu} = \nabla\varphi_{\#}\mu$. So (b) implies (a), which completes the proof. □

3. A monotonicity theorem

In this section we still admit the main theorem of this paper, Theorem 1.2. We generalize the notion of c-monotonicity, which plays an important role in optimal transport theory [61, 26], to cost functions c defined on $\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)$ and we use it to formulate a necessary condition of optimality for transport plans. Then we give a new proof of Strassen theorem on the existence of submartingale couplings with given marginals in dimension 1 using this c-monotonicity criterium.

3.1. A necessary condition of optimality.

The following definition generalizes the notion of c-monotonicity ; we refer to [59, 72, 73] and the references therein for a complete account on the subject. In what follows, we denote by $\mathcal{P}_1(\mathbb{R}^d)$ the set of probability measures on $\mathbb{R}^d$ having finite first moment.

Definition 3.1. *Let $c : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \to \mathbb{R}^+$ be a cost function. We will say that a set $\Gamma \subset \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)$ is c-monotone if for any $N \geq 1$ and points $(x_i, p_i) \in \Gamma$ and probability*

measures $q_i \in \mathcal{P}_1(\mathbb{R}^d)$, $i \in \{1, \dots, N\}$ *such that* $\sum_{i=1}^N p_i = \sum_{i=1}^N q_i$, *it holds*

$$\sum_{i=1}^N c(x_i, p_i) \leq \sum_{i=1}^N c(x_i, q_i).$$

We now formulate a necessary condition for optimality of a coupling for the transport cost $\overline{\mathcal{T}}_2$ which we recall is associated to the cost function c_2 defined by

$$c_2(x, p) = \left| \int y \, dp(y) - x \right|^2, \qquad x \in \mathbb{R}^d, \qquad p \in \mathcal{P}_1(\mathbb{R}^d).$$

Theorem 3.1. *Let* $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$. *There exists a set* $\Gamma \subset \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d)$ *which is monotone with respect to the cost function* c_2 *and such that if* $d\pi(x, y) = d\mu(x) dp_x(y)$ *is an optimal coupling for the cost* $\overline{\mathcal{T}}_2(\nu|\mu)$, *it holds*

$$\mu\left(\{x \in \mathbb{R}^d : (x, p_x) \in \Gamma\}\right) = 1.$$

Remark 3.1. *The paper* [5] *contains a general version of Theorem 3.1 relying on a completely different proof (see* [5, Theorem 3.4]*). It is also shown in* [5, Theorem 3.6] *that, under mild assumptions on the cost function (verified for instance by the cost* c_2*),* $c-$*monotonicity is also a sufficient condition of optimality.*

Proof. Recall the duality formula

$$\overline{\mathcal{T}}_2(\nu|\mu) = \sup_f \left\{ \int Q_2 f(x) \, d\mu(x) - \int f(y) \, d\nu(y) \right\},$$

where the infimum is running over the set of convex functions bounded from below and $Q_2 f(x) = \inf_{y \in \mathbb{R}^d} \{f(y) + |y - x|^2\}$, $x \in \mathbb{R}^d$. Note that if f is convex, then

$$Q_2 f(x) = \inf_{p \in \mathcal{P}_1(\mathbb{R}^d)} \left\{ c_2(x, p) + \int f \, dp \right\}, \qquad x \in \mathbb{R}^d.$$

Indeed, the inequality $\geq$ is obtained by taking $p = \delta_y$. Moreover, since f is convex, it holds

$$\int f(y) \, dp(y) + c_2(x, p) \geq f\left(\int y \, dp(y) \right) + \left| \int y \, dp(y) - x \right|^2 \geq Q_2 f(x)$$

so taking the infimum over p gives the converse inequality. Let f° be such that $\overline{\mathcal{T}}_2(\nu|\mu) = \int Q_2 f^\circ(x) \, dp(x) - \int f^\circ(y) \, d\nu(y)$ (which exists according to Theorem 1.2). Then, it holds

$$\begin{aligned}
\overline{\mathcal{T}}_2(\nu|\mu) &= \int Q_2 f^\circ(x) \, d\mu(x) - \int f^\circ(y) \, d\nu(y) \\
&\leq \int \left(\int f^\circ(y) \, dp_x(y) + c_2(x, p_x) \right) d\mu(x) - \int f^\circ(y) \, d\nu(y) \\
&= \int c_2(x, p_x) \, d\mu(x) = \overline{\mathcal{T}}_2(\nu|\mu).
\end{aligned}$$

Therefore, $Q_2 f^\circ(x) = \int f^\circ(y) \, dp_x(y) + c_2(x, p_x)$ for μ-almost all $x \in \mathbb{R}^d$. So denoting by Γ the set of couples (x, p) so that $Q f^\circ(x) = \int f^\circ \, dp + c_2(x, p)$, we have that $\mu(\{x \in \mathbb{R}^d : (x, p_x) \in \Gamma\}) = 1$. Now, let us check that Γ is monotone with respect to the cost function c_2. Take a family of points $(x_i, p_i) \in \Gamma$ and probability measures $q_i \in \mathcal{P}_1(\mathbb{R}^d)$, $i \in \{1, \dots, N\}$ such that $\sum_{i=1}^N p_i = \sum_{i=1}^N q_i$. Then, it holds

$$Q_2 f^\circ(x_i) = c_2(x_i, p_i) + \int f^\circ \, dp_i, \qquad \forall i \in \{1, \dots, N\}$$

and on the other hand

$$Q_2 f^\circ(x_i) \leq c_2(x_i, q_i) + \int f^\circ \, dq_i, \qquad \forall i \in \{1, \dots, N\}.$$

Summing these inequalities, and using the fact that $\sum_{i=1}^N p_i = \sum_{i=1}^N q_i$ gives immediately $\sum_{i=1}^N c(x_i, p_i) \leq \sum_{i=1}^N c(x_i, q_i)$. □

Remark 3.2. *Using the notation of Theorem 1.2, let us show how, with Theorem 3.1, one can recover the fact proved in Theorem 1.2 that $\nabla\varphi$ is 1-Lipschitz on the support of μ. Let $d\pi(x,y) = d\mu(x)dp_x(y)$ be an optimal transport plan from μ to ν and Γ be a c_2-monotone set such that $(x, p_x) \in \Gamma$ for μ almost every $x \in \mathbb{R}^d$. If (x, p_x) and (y, p_y) are elements of Γ, then comparing this pair with $(x, (1-\varepsilon)p_x + \varepsilon p_y)$ and $(y, (1-\varepsilon)p_y + \varepsilon p_x)$ where $\varepsilon \in [0,1]$ yields*

$$|x - \bar{x}|^2 + |y - \bar{y}|^2 \leq |x - [(1-\varepsilon)\bar{x} + \varepsilon\bar{y}]|^2 + |y - [(1-\varepsilon)\bar{y} + \varepsilon\bar{x}]|^2,$$

where $\bar{x} = \int z\, dp_x(z)$ and $\bar{y} = \int z\, dp_z(z)$. Noting that this is an equality for $\varepsilon = 0$ and taking the right derivative at $\varepsilon = 0_+$, we thus obtain

$$2\langle \bar{x} - x, \bar{y} - \bar{x}\rangle - 2\langle \bar{y} - y, \bar{y} - \bar{x}\rangle = 2\langle y - x, \bar{y} - \bar{x}\rangle - 2|\bar{y} - \bar{x}|^2 \geq 0,$$

hence $|\bar{y} - \bar{x}| \leq |y - x|$. Since $\bar{x} = \nabla\varphi(x)$ and $\bar{y} = \nabla\varphi(y)$ this proves that $\nabla\varphi$ is 1-Lipschitz on the support of μ.

3.2. **An example related to the increasing convex order.**

Definition 3.2. *Let μ, ν be two probability measures on $\mathbb{R}$; μ is dominated by ν for the* stochastic order*, denoted by $\mu \preceq_{\text{sto}} \nu$, if $\int f\, d\mu \leq \int f\, d\nu$ for all bounded increasing functions $f : \mathbb{R} \to \mathbb{R}$. Assuming that μ, ν have finite first moments, μ is dominated by ν for the* increasing convex order*, denoted by $\mu \preceq_{c,\text{sto}} \nu$, if $\int f\, d\mu \leq \int f\, d\nu$ for every increasing convex function $f : \mathbb{R} \to \mathbb{R}$.*

Notice that if $\int x\, d\mu(x) = \int y\, d\nu(y)$ and $\mu \preceq_{c,\text{sto}} \nu$, then it is easy to see that the inequality $\int f\, d\mu \leq \int f\, d\nu$ can be extended, first to all convex functions f with finite asymptotic slopes at $-\infty$, and second, by using the monotone convergence theorem, to all convex functions. Therefore, for probability measures with finite first moments, $\mu \preceq_c \nu$ is equivalent to "$\mu \preceq_{c,\text{sto}} \nu$ and $\int x\, d\mu(x) = \int x\, d\nu(x)$".

In this section we prove that on $\mathbb{R}$ under the assumption $\mu \preceq_{c,\text{sto}} \nu$, our Theorem 1.2 on the quadratic barycentric transport problem can be completed with new informations.

Theorem 3.2. *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$ be such that $\mu \preceq_{c,\text{sto}} \nu$. Then, with the notations of Proposition 1.1 and Theorem 1.2, the probability measure $\bar{\mu}$ satisfies $\mu \preceq_{\text{sto}} \bar{\mu}$ and any optimal coupling (X, Y) satisfies $X \leq \mathbb{E}[Y|X]$ almost surely.*

In the terminology of martingale transport, (X, Y) is a submartingale coupling and its law a submartingale transport plan.

Before we prove Theorem 3.2, note that conversely to it, if (X, Y) is a coupling of μ and ν such that $X \leq \mathbb{E}[Y|X]$ almost surely, then for every increasing and convex function $f : \mathbb{R} \to \mathbb{R}$, it holds

$$\mathbb{E}[f(Y)] \geq \mathbb{E}[\mathbb{E}[f(Y)|X]] \geq \mathbb{E}[f(\mathbb{E}[Y|X])] \geq \mathbb{E}[f(X)].$$

and so $\mu \preceq_{c,\text{sto}} \nu$. We therefore recover the classical result by Strassen [69, Theorem 9] on submartingale couplings (under the additional assumption that the measures have finite second moments): $\mu \preceq_{c,\text{sto}} \nu$ if and only if there exists $X \sim \mu$ and $Y \sim \nu$ in the same probability space with $X \leq \mathbb{E}[Y|X]$ almost surely.

Remark 3.3. *Theorem 3.2 also gives back Strassen Theorem in the very specific case of the real line and measures with finite second moments. Let us prove both implications. The fact that $\mathbb{E}[Y|X] = X$ implies $\mu \preceq_c \nu$ is an obvious consequence of the conditional Jensen inequality. Conversely, if $\mu \preceq_c \nu$ first notice $\mu \preceq_{c,\text{sto}} \nu$. Therefore according to Theorem 3.2 there exists $X \sim \mu$ and $Y \sim \nu$ in the same probability space with $\mathbb{E}[Y|X] \geq X$. Moreover, $\mu \preceq_c \nu$ implies $\int x\, d\mu(x) = \int x\, d\nu(x)$, i.e. $\mathbb{E}[X] = \mathbb{E}[Y]$. Hence $\mathbb{E}[Y|X]$ and X are ordered random variables with the same expectation. They are therefore almost surely*

equal, which proves the converse implication of Strassen Theorem. For another elementary proof, see [7, Section 2]

Proof of Theorem 3.2. We denote a solution of the quadratic barycentric problem by $d\pi(x,y) = d\mu(x)dp_x(y)$. Let (X,Y) such that $(X,Y) \sim \pi$ and denote $\mathbb{E}[Y|X]$ by $\bar{X}$. According to Theorem 1.2, π may not be uniquely determined but the law of $\bar{X}$ is uniquely determined and denoted by $\bar{\mu}$. The inequality $X \leq \mathbb{E}[Y|X]$ almost surely immediately implies that $\mu \preceq_{\text{sto}} \bar{\mu}$. Now to show that $X \leq \mathbb{E}[Y|X]$ almost surely amounts to show that $\int y\, dp_x(y) \geq x$ for μ almost all $x \in \mathbb{R}$. The rest of the proof is devoted to this question.

According to Theorem 3.1, there exists $\Gamma \subset \mathbb{R} \times \mathcal{P}_1(\mathbb{R})$ such that, $\mu(\{x \in \mathbb{R} : (x,p_x) \in \Gamma\}) = 1$ and such that if $(x,p) \in \Gamma$ and $(x',p') \in \Gamma$, then for all probability measures q, q' such that $p + p' = q + q'$, it holds

$$\left|\int y\, dp(y) - x\right|^2 + \left|\int y\, dp'(y) - x'\right|^2 \leq \left|\int y\, dq(y) - x\right|^2 + \left|\int y\, dq'(y) - x'\right|^2. \tag{6}$$

Let us show that if $x, x' \in \mathbb{R}$ are such that $(x,p_x) \in \Gamma$ and $(x',p_{x'}) \in \Gamma$, then for all $a \in \text{Support}(p_x)$ and $b \in \text{Support}(p_{x'})$, it holds

$$\left[\left(\int y\, dp_x(y) - x\right) - \left(\int y\, dp_{x'}(y) - x'\right)\right](b-a) \geq 0. \tag{7}$$

Let $a \in \text{Support}(p_x)$ and $b \in \text{Support}(p_{x'})$ and $\varepsilon > 0$ and define

$$r_a = p_x(\,\cdot\,|[a-\varepsilon, a+\varepsilon]) \qquad \text{and} \qquad r_b = p_{x'}(\,\cdot\,|[b-\varepsilon, b+\varepsilon]).$$

For all $t > 0$, define

$$q_x^t = p_x + t(r_b - r_a) \qquad \text{and} \qquad q_{x'}^t = p_{x'} + t(r_a - r_b)$$

and note that if t is small enough, q_x^t and $q_{x'}^t$ are probability measures such that $q_x^t + q_{x'}^t = p_x + p_{x'}$. Applying (6) and letting $t \to 0$ gives

$$\left[\left(\int y\, dp_x(y) - x\right) - \left(\int y\, dp_{x'}(y) - x'\right)\right]\left(\int y\, dr_b(y) - \int y\, dr_a(y)\right) \geq 0.$$

Finally, letting $\varepsilon \to 0$ yields to (7).

Define $\Gamma^1 = \{x \in \mathbb{R} : (x,p_x) \in \Gamma\}$ and for any interval $I \subset \mathbb{R}$, let A_I the set of points $x \in \Gamma^1$ such that $\int y\, dp_x(y) - x \in I$. We aim at proving that $\mu(A_{]-\infty,0[}) = 0$.

Striving for a contradiction suppose that $\mu(A_{]-\infty,0[}) > 0$. Our assumption $\mu \preceq_{c,\text{sto}} \nu$ implies $\int x\, d\mu(x) \leq \int y\, d\nu(y)$ so that $\int_{\Gamma^1}[(\int y\, dp_x(y)) - x]d\mu(x) \geq 0$. It follows that $\mu(A_{[0,+\infty[}) > 0$. Henceforth $A_{]-\infty,0[}$ and $A_{[0,+\infty[}$ are not empty. Several configurations may be considered.

- In the first case we assume that there exists $x' \in A_{]-\infty,0[}$ and $x \in A_{[0,+\infty[}$ such that $x' \leq x$. One thus has $\int y\, dp_{x'}(y) < x' \leq x \leq \int y\, dp_x(y)$. So in this case, one can find $a \in \text{Support}(p_x)$ and $b \in \text{Support}(p_{x'})$ such that $a > b$. Applying (7) with these a, b provides a contradiction.
- We can now assume that $A_{[0,+\infty[}$ has a supremum smaller than or equal to the infimum of $A_{]-\infty,0[}$. If there exists $a \in \text{Support}(p_x)$ and $b \in \text{Support}(p_{x'})$ such that $a > b$, we get the same contradiction as before.
- We are finally reduced to the case where $A_{[0,+\infty[}$ has a supremum smaller than or equal to the infimum of $A_{]-\infty,0[}$ and moreover every p_x with $x \in A_{[0,+\infty[}$ has the essential supremum smaller than or equal to the essential infimum of every $p_{x'}$ with $x' \in A_{]-\infty,0[}$. Let $m \in \mathbb{R}$ be between $\sup_{x \in A_{[0,+\infty[}}[\sup \text{Support}(p_x)]$ and $\inf_{x' \in A_{]-\infty,0[}}[\inf \text{Support}(p_{x'})]$. Integration of $x \mapsto (x-m)_+$ provides a contradiction to $\mu \preceq_{c,\text{sto}} \nu$. Indeed, one the one hand, it holds

$$\int [x-m]_+\, d\mu(x) \geq \int [x-m]_+ \mathbf{1}_{A_{]-\infty,0[}}(x)\, d\mu(x)$$

(with actually equality). On the other hand,

$$\begin{aligned}\int [y-m]_+ \, d\nu(y) &= \int \left(\int [y-m]_+ \, dp_x(y) \right) d\mu(x) \\ &\overset{(i)}{=} \int \left(\int [y-m]_+ \, dp_x(y) \right) \mathbf{1}_{A_{]-\infty,0[}}(x) \, d\mu(x) \\ &\overset{(ii)}{=} \int \left(\int y - m \, dp_x(y) \right) \mathbf{1}_{A_{]-\infty,0[}}(x) \, d\mu(x) \\ &\overset{(iii)}{<} \int (x-m) \, \mathbf{1}_{A_{]-\infty,0[}}(x) \, d\mu(x) \\ &\overset{(iv)}{=} \int [x-m]_+ \mathbf{1}_{A_{]-\infty,0[}}(x) \, d\mu(x),\end{aligned}$$

where

- (i) comes from the fact that for any $x \in A_{[0,+\infty[}$, $\mathrm{Support}(p_x) \subset]-\infty, m]$,
- (ii) comes from the fact that for any $x \in A_{]-\infty,0[}$, $\mathrm{Support}(p_x) \subset [m, +\infty[$,
- (iii) comes from the definition of $A_{]-\infty,0[}$ and the fact that if $f < g$ on a set A such that $\mu(A) > 0$, then $\int_A f \, d\mu < \int_A g \, d\mu$,
- (iv) follows from the inequality $x > \int y \, dp_x(y) \geq m$, for all $x \in A_{]-\infty,0[}$.

In any case there is a contradiction to the fact that $\mu(A_{]-\infty,0[}) > 0$ and this completes the proof. □

Remark 3.4.

(1) *Optimal transport plans for the weak transport problem are not uniquely determined. As stated in Theorem 1.2, $\bar{\mu}$ and the transport from μ to $\bar{\mu}$ are uniquely determined but the martingale transport plan from $\bar{\mu}$ to ν is completely free. It is precisely the goal of Martingale Optimal Transport to determine special martingale transport plans from $\bar{\mu}$ to ν.*

(2) *This remark holds in particular for the dimension 1. Namely, in Theorem 3.2 we did not defined a submartingale coupling but only the non-decreasing/increasing part of its Doob decomposition. In relation with this decomposition, the paper [57] proposes two supermartingale transport problems and describes their solutions, extending the theory on curtain couplings initiated in [7]. In particular the two optimal supermartingale couplings respectively corresponding to the two problems coincide with the curtain coupling when $\mu \preceq_c \nu$. However, transforming the supermartingale problem in submartingale problem, we stress that $\bar{X} = \mathbb{E}[Y|X]$ and its coupling with X is different from ours as we show in this example: if X is the uniform measure on $[-1, 0]$ and Y is uniform on $[0, 3]$, our coupling is $\mathbb{E}[Y|X] = \bar{X} = X + 2$ (it is a translation see Section 4). The first supermartingale coupling by Nutz and Stebegg (reversed in order to be a submartingale) gives $Y = \mathbb{E}[Y|X] = 3(X + 1)$. The second one provides $Y = \mathbb{E}[Y|X] = -3X$.*

4. Example of ν concentrated on the vertices of a simplex

In this section we prove the following theorem, mainly by geometric means.

Theorem 4.1. *Let μ be a compactly supported probability measure on $\mathbb{R}^d$ and ν an atomic measure whose support is a simplex $\{y_0, \dots, y_k\}$ of $\mathbb{R}^d$ with $k \leq d$. Denote the convex hull of $\{y_0, \dots, y_k\}$ by Δ. Then there exists $v \in \mathbb{R}^d$ such that the map T defined by*

$$T : \mathbb{R}^d \to \mathbb{R}^d : x \mapsto T(x) = \mathrm{proj}_\Delta(x + v)$$

is such that $\bar{\mu} = T_\# \mu$ (with the notations of Proposition 1.1 and Theorem 1.2), where proj_Δ denotes the orthogonal projection on the closed convex set Δ.

Remark 4.1.

- *To be consistent with Theorem 1.2, note that the map T given above can be written as $T = \nabla\varphi$, where $\varphi : \mathbb{R}^d \to \mathbb{R}$ is the convex function of class $\mathcal{C}^1$ defined by*
$$\varphi(x) = \frac{|x+v|^2}{2} - \frac{1}{2}d(x+v,\Delta)^2, \qquad x \in \mathbb{R}^d,$$
where $d(z,\Delta) = \inf_{y\in\Delta}|z-y|$.
- *The assumption that μ is compactly supported could be easily relaxed into the assumption that μ admits a moment of order 2 finite.*

Proof. Let μ be a probability with compact support and ν an atomic measure with support the simplex $\{y_0,\dots,y_k\}$ of $\mathbb{R}^d$ with $k \le d$. We denote by $\bar\mu$ the projection of μ on $B_\nu = \{\eta \in \mathcal{P}(\mathbb{R}^d) : \eta \preceq_c \nu\}$ given by Proposition 1.1 and Theorem 1.2. The assumption that the support of ν is a simplex is in order for the following property: for any point y in the convex hull $\Delta = \mathrm{Conv}(y_0,\dots,y_k)$ the barycentric coordinates $(\lambda_0,\dots,\lambda_k)$, defined by $\sum_{i=0}^k \lambda_i = 1$ and $\sum_{i=0}^k \lambda_i y_i = y$, are uniquely determined. Moreover, notice that all the coordinates are non-negative. If $v \in \mathbb{R}^d$, we will denote by μ_v the translation of μ by the vector v, *i.e* $\mu_v = \mathrm{Law}(X+v)$, with $X \sim \mu$. The four following properties will permit us to determine $\bar\mu$ and find the optimal coupling between μ and $\bar\mu$.

(a) Let η be a probability measure with a finite moment of order 2. A coupling (X,Y) is an optimal coupling for W_2 between μ and η if and only if $(X+v,Y)$ is an optimal coupling for W_2 between μ_v and η. Moreover,
$$W_2^2(\mu_v,\eta) = W_2^2(\mu,\eta) + |v|^2 + 2\left\langle \int x\,d\mu - \int x\,d\eta, v\right\rangle,$$
(b) As $\bar\mu \preceq_c \nu$ the measure $\bar\mu$ is concentrated on Δ and has the same barycenter as ν. Conversely any measure η concentrated on Δ and with the same barycenter as ν satisfies $\eta \preceq_c \nu$.
(c) Among the measures concentrated on Δ, the one that minimises the quadratic cost W_2 with respect to a given probability measure μ' having a finite moment of order 2 is $(\mathrm{proj}_\Delta)_\#\mu'$.
(d) There exists $v \in \mathbb{R}^d$ such that $(\mathrm{proj}_\Delta)_\#\mu_v$ has the same barycenter as ν.

Before we prove the different four points, let us finish the proof. According to Item (b), all the elements of B_ν have the same barycenter as ν. Therefore, applying Item (a), we have for any $\eta \in B_\nu$
$$W_2^2(\mu_v,\eta) = W_2^2(\mu,\eta) + |v|^2 + 2\left\langle \int x\,d\mu - \int x\,d\nu, v\right\rangle.$$
One easily concludes from this identity that $\overline{\mu_v} = \bar\mu$ for any $v \in \mathbb{R}^d$. According to Item (d), there exists $v \in \mathbb{R}^d$ such that $\eta_v := (\mathrm{proj}_\Delta)_\#\mu_v$ has the same barycenter as ν. Since η_v is also concentrated on Δ, it follows from Item (b) that η_v belongs to B_ν. Therefore, according to Item (c),
$$\inf_{\eta\in B_\nu} W_2^2(\mu_v,\eta) \ge \inf_{\eta(\Delta)=1} W_2^2(\mu_v,\eta) = W_2^2(\mu_v,\eta_v)$$
and so $\eta_v = \overline{\mu_v} = \bar\mu$. Finally if $X \sim \mu$, it follows from Item (a) that $(X, \mathrm{proj}_\Delta(X+v))$ is an optimal coupling between μ and $\bar\mu$.

The four points above can be proved as follows:

(a) This assumption simply comes from
$$\mathbb{E}[|(X+v)-\bar X|^2] = \mathbb{E}[|X-\bar X|^2] + \underbrace{|v|^2 + 2\langle \mathbb{E}[X-\bar X], v\rangle}_{\text{Depends only on } \mathrm{Law}(X),\,\mathrm{Law}(\bar X)}.$$
(b) The first implication is obvious. For the second implication, assume η is concentrated on Δ and has the same barycenter as ν. For every $x \in \Delta$, let p_x be the unique probability measure concentrated on $\{y_0,\dots,y_k\}$ with barycenter x and let

$\nu' = \int p_x \, d\eta(x)$. The probability measure ν' is concentrated on the same set as ν. Its barycenter permits to determine it uniquely. The barycenter of ν' is $\int x \, d\eta(x)$, the barycenter of η that is also the one of ν. Therefore $\nu' = \nu$. We have proved that there exists a martingale having η on ν as marginals. Therefore $\eta \preceq_c \nu$.

(c) For every $x \in \mathbb{R}^d$, the point $\mathrm{proj}_\Delta(x)$ is by definition the closest point in Δ. So if $X' \sim \mu'$ and Y takes values in Δ almost surely, one gets

$$\mathbb{E}[|X' - Y|^2] \geq \mathbb{E}[|X' - \mathrm{proj}_\Delta(X')|^2]$$

and so $W_2^2(\mu', \eta) \geq \mathbb{E}[|X' - \mathrm{proj}_\Delta(X')|^2]$ for any η concentrated in Δ. In particular, $W_2^2(\mu', (\mathrm{proj}_\Delta)_\# \mu') = \mathbb{E}[|X' - \mathrm{proj}_\Delta(X')|^2]$ which proves the claim.

(d) Without loss of generality we can assume $k = d$ for the following reason: if Δ has positive codimension, the measure $(\mathrm{proj}_\Delta)_\# \mu'$ is exactly $(\mathrm{proj}_\Delta)_\# \circ (\mathrm{proj}_A)_\# \mu'$ where A is the affine space spanned by Δ. As a consequence, in our investigation we can replace μ by $(\mathrm{proj}_A)_\# \mu$, consider only translations in A and see Δ as a simplex with full dimension.

Now, let y_ν be the barycenter of ν and let us prove the existence of $v \in \mathbb{R}^d$ such that the barycenter of $(\mathrm{proj}_\Delta)_\# \mu_v$ is equal to y_ν. This point can be proved using the notion of topological degree coming from algebraic topology (we refer to [58, Chapter IV] for an introduction). If Ω is a bounded open set of $\mathbb{R}^d$ and $f : \overline{\Omega} \to \mathbb{R}^d$ is a continuous mapping, the degree of f is a $\mathbb{Z}$-valued quantity denoted by $\mathsf{d}(f, \Omega, a)$ defined for every $a \in \mathbb{R}^d \setminus f(\partial\Omega)$. We will use the following basic properties:

- if $f = \mathrm{Id}$, then $\mathsf{d}(\mathrm{Id}, \Omega, a) = 1$, for all $a \in \Omega$;
- if $\mathsf{d}(f, \Omega, a) \neq 0$, then the equation $f(x) = a$, $x \in \Omega$, admits at least one solution;
- if $F : \overline{\Omega} \times [0,1] \to \mathbb{R}^d$ is a continuous function and $a \in \mathbb{R}^d$ is such that for all $t \in [0,1]$, $a \in \mathbb{R}^d \setminus F(t, \partial\Omega)$, then $\mathsf{d}(F(\,\cdot\,, 0), \Omega, a)) = \mathsf{d}(F(\,\cdot\,, 1), \Omega, a))$ (invariance by homotopy).

Consider X a random variable with law μ and $R > 0$ a positive number to be fixed later such that the open ball of center 0 and radius R denoted by $\mathcal{B}(0, R)$ contains Δ. From now on the sphere of center 0 and radius R will be denoted by $\mathcal{S}(0, R)$. We are interested in the map

$$\Phi : \overline{\mathcal{B}}(0, R) \times [0,1] \to \Delta \subset \mathbb{R}^d, \tag{8}$$

defined so that $\Phi(v, t)$ is the barycenter of $(\mathrm{proj}_\Delta)_\#(\mathrm{Law}(v + tX))$. Note that this map is continuous so that it can be seen as an homotopy between proj_Δ and $v \mapsto \int y \, d(\mathrm{proj}_\Delta)_\# \mu_v(y)$. Another homotopy is possible between Id and proj_Δ

$$\Psi : \overline{\mathcal{B}}(0, R) \times [0,1] \to \mathbb{R}^d$$

defined by $\Psi(x, t) = (1 - t)x + t\mathrm{proj}_\Delta(x)$. Note that since y_ν lies in the interior of Δ, it does not belong to $\Psi(\mathcal{S}(0, R) \times [0,1])$. Therefore, by invariance by homotopy,

$$1 = \mathsf{d}(\mathrm{Id}, \mathcal{B}(0, R), y_\nu) = \mathsf{d}(\Psi(\cdot, 0), \mathcal{B}(0, R), y_\nu) = \mathsf{d}(\Psi(\cdot, 1), \mathcal{B}(0, R), y_\nu),$$

from which we infer that $\mathsf{d}(\Phi(\cdot, 0), \mathcal{B}(0, R), y_\nu) = 1$. According to Lemma 4.1 below, if R is large enough then $\Phi(\mathcal{S}(0, R) \times [0,1]) \subset \partial\Delta$. Since y_ν lies in the interior of Δ, we conclude that $\Phi(\mathcal{S}(0, R) \times [0,1])$ does not meet y_ν, and so using the homotopy invariance again, we get that $\mathsf{d}(\Phi(\cdot, 1), \mathcal{B}(0, R), y_\nu) = 1$ and so the equation $\Phi(v, 1) = \int \mathrm{proj}_\Delta(x) \, d\mu_v(x) = y_\nu$ admits at least one solution $v \in \mathcal{B}(0, R)$ which completes the proof. □

Lemma 4.1. *Under the preceding assumptions, if R is large enough, the map Φ defined in (8) is such that $\Phi(\mathcal{S}(0, R) \times [0,1]) \subset \partial\Delta$.*

Proof. Recall that

$$\Delta = \left\{ \sum_{i=0}^{d} \lambda_i y_i : \sum_{i=0}^{d} \lambda_i = 1,\ \lambda_0, \dots, \lambda_d \geq 0 \right\}.$$

For J a subset of $\{0, \dots, d\}$ we denote by Δ_J the set

$$\Delta_J = \left\{ \sum_{i=0}^{d} \lambda_i y_i : \forall i \in J,\ \lambda_i = 0 \right\} \subset \Delta.$$

We denote by u_j the outward normal unit vector of the facet $\Delta_{\{j\}}$. Recall that the relative interior of a subset is the interior of this set in the topology induced by its affine span. Every point of $\mathbb{R}^d$ whose projection belongs to the relative interior of Δ_J can be written in the form $\sum_{i \notin J} \lambda_i y_i + \sum_{i \in J} \xi_i u_i$ where the coefficients $(\lambda_i)_{i \notin J}$ are positive and satisfy $\sum_{i \notin J} \lambda_i = 1$ and the coefficients $(\xi_i)_{i \in J}$ are non-negative.

We prove now that for every $r > 0$, there exists $R_0 > 0$ such that for all $R \geq R_0$ it holds

- Δ is contained in $\mathcal{B}(0, R - r + 1)$,
- for every $v \in \mathcal{S}(0, R)$ there exists $j \in \{0, \dots, d\}$ such that $\mathrm{proj}_\Delta(\mathcal{B}(v, r)) \subset \Delta_{\{j\}}$.

Therefore if $X \sim \mu$ with μ such that $\mu(\mathcal{B}(0, r)) = 1$ and $v \in \mathcal{S}(0, R)$, there exists j such that the barycenter of $(\mathrm{proj}_\Delta)_\#(\mathrm{Law}(v + tX))$ is in $\Delta_{\{j\}} \subset \partial\Delta$, which proves the claim.

Let us consider

$$g : x \in \mathbb{R}^d \setminus \Delta \mapsto (x - \mathrm{proj}_\Delta(x))/|x - \mathrm{proj}_\Delta(x)| \in \mathbb{S}^{d-1}.$$

For our proof it is enough to find R such that for every $x \in \mathcal{S}(0, R)$ the range $g(\mathcal{B}(x, r))$ only contains vectors of $\mathbb{S}^{d-1}$ that are in the cones $(C_J)_{J \ni j, \#J \leq d}$ for some j, where

$$C_J = \left\{ u \in \mathbb{R}^d : u = \sum_{i \in J} \xi_i u_i,\ \xi_i \geq 0,\ \forall i \in J \right\}.$$

In particular it does not intersect $C_{\{0,\dots,d\} \setminus \{j\}}$. Striving for a contradiction we assume that there exists an increasing sequence $R_n \to \infty$ and $x_n \in \mathcal{S}(0, R_n)$ such that the property is not satisfied. Therefore, for every n the sets $g(\mathcal{B}(x_n, r)) \subset \mathbb{S}^{d-1}$ intersects for every j a cone C_J with $j \in J$. However, the diameter of $g(\mathcal{B}(x_n, r))$ tends to zero. Up to selecting a subsequence, the sequence converges in the Hausdorff topology to a compact set of diameter zero, i.e a point $\{u_\infty\} \subset \mathbb{S}^{d-1}$. We have $u_\infty \in C_{J_\infty}$ for some $J_\infty \subset \{0, \dots, d\}$. Let j_∞ be in J_∞. We have also $g(x_n) \to u_\infty$ (up to a subsequence) and the fact that g is locally Lipschitz with a constant that tends to zero at infinity tells us that the j_∞ coordinate is not zero in the cone coordinate $u = \sum_j \xi_j u_j$ for all the points u of $g(\mathcal{B}(x_n, r)) \subset \mathbb{S}^{d-1}$ when n is great enough - a contradiction. □

5. Other examples and discussion of the literature

In this section, we briefly present and discuss other examples of optimal transport problems involving weak cost functions.

(1) In [31, Theorem 1.5], it is shown that if μ, ν are probability measures on $\mathbb{R}$ having finite first moments and if $\bar{\mu} \in B_\nu$ denotes the projection of μ on B_ν as defined in Proposition 1.1, then for any even convex cost function $\theta : \mathbb{R} \to \mathbb{R}_+$, it holds

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \inf_{\pi \in C(\mu, \bar{\mu})} \iint \theta(y - x)\, d\pi(x, y),$$

where we recall that the barycentric optimal transport cost $\overline{\mathcal{T}}_\theta(\nu|\mu)$ is defined by (3). This result has been recently recovered and completed by Alfonsi, Corbetta and Jourdain in [1], where an explicit expression is given for $\bar{\mu}$ (see [1, Proposition 3.4]).

(2) In the recent paper [2] (see in particular the section 5.2 of [2]), Alibert, Bouchitté and Champion consider the family of cost functions $(c_\lambda)_{\lambda\geq 0}$ defined for all $x \in \mathbb{R}^d$ and $p \in \mathcal{P}(\mathbb{R}^d)$ having a finite first moment by

$$c_\lambda(x,p) = (\lambda-1)\int |y-x|^2\,dp(y) + \left|\int y\,dp(y) - x\right|^2$$

and study the associated optimal transport problem:

$$\overline{\mathcal{T}}_{c_\lambda}(\nu|\mu) = \inf_{\pi\in C(\mu,\nu)} \int c_\lambda(x,p_x)\,d\mu(x)$$

(this quantity is denoted by $F_\lambda(\mu,\nu)$ in [2]). It turns out that our Theorem 1.2 yields a full description of optimal couplings for these costs for $\lambda > 0$ (which completes the somehow implicit characterization of [2, Theorem 5.6]). Namely, an easy calculation reveals that if $d\pi(x,y) = d\mu(x)dp_x(y) \in C(\mu,\nu)$ where μ,ν are compactly supported, then

$$\int c_\lambda(x,p_x)\,d\mu(x) = C(\lambda) + \int \left|\int y\,dp_{x/\lambda}(y) - x\right|^2 d\mu_\lambda(x),$$

where $C(\lambda) = -\lambda(\lambda-1)\int |x|^2\,d\mu(x) + (\lambda-1)\int |y|^2\,d\nu(y)$ and μ_λ is the image of μ under the map $x \mapsto \lambda x$. Since the kernel $q_x = p_{x/\lambda}$, $x \in \mathbb{R}^d$, is such that $\nu(\,\cdot\,) = \int q_x(\,\cdot\,)\,d\mu_\lambda(x)$, we conclude that

$$\overline{\mathcal{T}}_{c_\lambda}(\nu|\mu) = C(\lambda) + \overline{\mathcal{T}}_2(\nu|\mu_\lambda),$$

and p is optimal in $\overline{\mathcal{T}}_{c_\lambda}(\nu|\mu)$ if and only if q is optimal in $\overline{\mathcal{T}}_2(\nu|\mu_\lambda)$. Moreover, Theorem 1.2 yields that optimal kernels p are all of the form

$$p_x = r_{\nabla\varphi_\lambda(\lambda x)}, \quad \text{with } r \text{ such that } \int y\,dr_u(y) = u, \text{ for } \overline{\mu_\lambda} \text{ almost all } u \in \mathbb{R}^d,$$

where φ_λ is the $\mathcal{C}^1$ smooth convex function associated to the transport from μ_λ to the projection $\overline{\mu_\lambda}$ of μ_λ on B_ν. In terms of random vectors, $\pi_\lambda = \mathrm{Law}(X_\lambda, Y_\lambda) \in C(\mu,\nu)$ is optimal for c_λ if and only if $\mathbb{E}[Y_\lambda|X_\lambda] = \nabla\varphi_\lambda(\lambda X_\lambda)$ (which has distribution $\overline{\mu_\lambda}$).

(3) The case $\lambda = 0$ is also interesting since, in this case,

$$\overline{\mathcal{T}}_{c_0}(\nu|\mu) = -\sup_{\pi\in C(\mu,\nu)} \int \mathrm{Var}(p_x)\,d\mu(x),$$

where as usual $d\pi(x,y) = d\mu(x)dp_x(y)$ and for all $p \in \mathcal{P}(\mathbb{R}^d)$ having finite first moment, $\mathrm{Var}(p) = \int |y|^2\,dp(y) - |\int y\,dp(y)|^2$ (note that this is a concave function of p). In this case, as observed in [2], the unique optimal coupling π° is the product one : $\pi^\circ = \mu\otimes\nu$.

(4) As proved in [8, Proposition 5.2], the so-called *shadow couplings* are solutions to a weak optimal transport problem. Shadow couplings from μ to ν are martingale transport parametrised by a measure $\hat{\mu}$ with marginals λ and μ. Let $(\lambda_x)_{x\in[0,1]}$ be a disintegration of $\hat{\mu}$ with respect to μ. Then

$$c^{\hat{\mu}}(x,p_x) = \inf \int (1-u)\sqrt{1+y^2}\,d\alpha(u,y)$$

where the inf goes among all α with first marginal λ_x and second marginal p_x and such that we have $\int f(u)(y-x)\,d\alpha(u,y) = 0$ for every bounded f.

(5) Last but not least, let us mention that after the completion of this work, we learned that a recent remarkable economics paper on optimal mechanisms for the multiple-good monopoly problem shows interesting similarities with our context. In [17], Daskalakis, Deckelbaum and Tzamos study the maximization problem among convex coordinate-wise nondecreasing 1-Lipschitz potential functions defined on some

d-dimensional rectangle. In their Theorem 2, they prove the strong duality between this problem and "strong" dual problem (for us a primal problem). This problem is a transport problem with ℓ_1-norm as cost function and the possibility to replace μ and ν by μ' and ν' with $\mu \preceq_{c,\text{sto}} \mu'$ and $\nu' \preceq_{c,\text{sto}} \nu$.

6. Proofs

This section contains the proofs of Proposition 1.1 and Theorem 1.2 and of the technical Lemma 2.1.

6.1. **Proof of the existence and uniqueness of the projection.** In what follows, we will equip $\mathcal{P}_2(\mathbb{R}^d)$ with the topology generated by the metric W_2. We recall (see *e.g.* [72, Theorem 7.12]) that if $(\eta_k)_{k\geq 1}$ is a sequence of elements of $\mathcal{P}_2(\mathbb{R}^d)$ and $\eta \in \mathcal{P}_2(\mathbb{R}^d)$, then $W_2(\eta_k, \eta) \to 0$ if and only if $\int f\, d\eta_k \to \int f\, d\eta$ for any continuous function f satisfying $|f|(x) \leq a + b|x|^2$, $x \in \mathbb{R}^d$, for some $a, b \geq 0$.

Proof of Proposition 1.1. Let us show that the set B_ν is compact in $\mathcal{P}_2(\mathbb{R}^d)$ for the topology induced by W_2.

First, let us show that B_ν is closed. Namely, B_ν can be written as follows

$$B_\nu = \left\{ \eta \in \mathcal{P}_2(\mathbb{R}^d) : \int f\, d\eta \leq \int f\, d\nu, \forall f : \mathbb{R}^d \to \mathbb{R} \text{ convex and Lipschitz} \right\}.$$

(The fact that one can restrict to convex and Lipschitz functions in the definition of the convex order is classical ; one can see this by noting that if f is convex then $f_n : x \mapsto \inf_{y\in\mathbb{R}^d}\{f(y)+n|x-y|\}$, $x \in \mathbb{R}^d$, is a sequence of n-Lipschitz convex functions converging to f monotonically.) Since functionals $\eta \mapsto \int f\, d\eta$ with f convex and Lipschitz are continuous for the W_2 topology, it follows that B_ν is closed.

Now let us show that B_ν is precompact. According to a variant of Prokhorov theorem (see *e.g.* [33, Theorem 9.10]), this is equivalent to show that

$$\sup_{\eta\in B_\nu} \int |x|^2 \mathbf{1}_{|x|>R}\, d\eta(x) \to 0$$

as R goes to ∞. Note that

$$|x|^2 \mathbf{1}_{|x|>R} \leq \max(0; 2|x| - R)^2, \qquad \forall x \in \mathbb{R}^d.$$

The function on the right hand side being convex, one thus gets

$$\sup_{\eta\in B_\nu} \int |x|^2 \mathbf{1}_{|x|>R}\, d\eta(x) \leq \int \max(0; 2|x| - R)^2\, d\nu(x) := \varepsilon(R).$$

The monotone convergence theorem then implies that $\varepsilon(R) \to 0$ as $R \to +\infty$, which shows that B_ν is precompact.

It follows from what precedes that B_ν is compact. The map $\eta \mapsto W_2(\eta, \mu)$ therefore reaches its minimum on B_ν.

Now let us prove that the minimizer is unique. This will follow from the strict convexity of $W_2^2(\mu, \cdot)$ along generalized geodesics with base point μ and from the convexity of the set B_ν along those generalized geodesics. More precisely, suppose that η_0, η_1 are in B_ν and let (X, Y_0, Y_1) be a random vector such that $X \sim \mu$, $Y_0 \sim \eta_0$, $Y_1 \sim \eta_1$ and so that $\text{Law}(X, Y_i) \in C(\mu, \eta_i)$ and $W_2^2(\mu, \eta_i) = \mathbb{E}[|X - Y_i|^2]$, $i = 0, 1$. Then define $\eta_t = \text{Law}((1-t)Y_0 + tY_1)$ for all $t \in [0, 1]$. First note that $\eta_t \in B_\nu$ for all $t \in [0, 1]$. Indeed, if f is a convex function on $\mathbb{R}^d$, it holds

$$\int f\, d\eta_t = \mathbb{E}[f((1-t)Y_0 + tY_1)] \leq (1-t)\mathbb{E}[f(Y_0)] + t\mathbb{E}[f(Y_1)] \leq \int f\, d\nu.$$

Moreover, according to Lemma 9.2.1 of [3], it holds

$$W_2^2(\mu, \eta_t) \leq (1-t)W_2^2(\mu, \eta_0) + tW_2^2(\mu, \eta_1) - t(1-t)W_2^2(\eta_0, \eta_1), \qquad \forall t \in [0, 1]$$

So if $W_2^2(\mu,\eta_0) = W_2^2(\mu,\eta_1) = \inf_{\eta\in B_\nu} W_2^2(\mu,\eta)$, then necessarily $\eta_0 = \eta_1$, which shows uniqueness.

Finally, let us show that $\overline{\mathcal{T}}_2(\nu|\mu) = W_2^2(\bar\mu,\mu)$. First, let (X,Y) be a coupling of μ and ν and set $X' = \mathbb{E}[Y|X]$. Since (X',Y) is a martingale, it follows that $\mathrm{Law}(X') \in B_\nu$. Therefore $\mathbb{E}[|X-X'|^2] \geq W_2^2(\mu,\bar\mu)$ and so optimizing gives $\overline{\mathcal{T}_2}(\nu|\mu) \geq W_2^2(\mu,\bar\mu)$. On the other hand, let $\pi \in C(\mu,\eta)$ be a coupling between μ and some $\eta \in B_\nu$. Since $\eta \preceq_c \nu$, one can construct a Markov chain (X,X',Y) such that $(X,X') \sim \pi$ and (X',Y) is a martingale. Then it holds,

$$\begin{aligned}\mathbb{E}[|X-X'|^2] &= \mathbb{E}[|X-\mathbb{E}[Y|X,X']|^2]\\ &= \mathbb{E}\left[\mathbb{E}[[|X-\mathbb{E}[Y|X,X']|^2|X]\right]\\ &\geq \mathbb{E}[|X-\mathbb{E}[Y|X]|^2] \geq \overline{\mathcal{T}}_2(\nu|\mu),\end{aligned}$$

where the inequality comes from Jensen inequality for conditional expectation. Optimizing over X' and η gives that $\overline{\mathcal{T}_2}(\nu|\mu) \leq W_2^2(\bar\mu,\mu)$ and completes the proof. □

Remark 6.1. *Note that it is easy to conclude using similar arguments that there always exists a deterministic map T transporting μ on $\bar\mu$ such that $W_2^2(\mu,\bar\mu) = \int |x-T(x)|^2\,d\mu(x)$. Indeed, suppose that $(X,\bar X)$ is an optimal coupling for $W_2^2(\bar\mu,\mu)$ and consider $X' = \mathbb{E}[\bar X|X]$. Then, the conditional Jensen inequality gives*

$$\mathbb{E}[|X-X'|^2] = \mathbb{E}[|X-\mathbb{E}[\bar X|X]|^2] \leq \mathbb{E}[|X-\bar X|^2] = W_2^2(\bar\mu,\mu). \tag{9}$$

Let $\eta' := \mathrm{Law}(X')$. Then $\eta' \preceq_c \bar\mu$ and $\bar\mu \preceq_c \nu$ so $\eta' \in B_\nu$. Therefore, $\mathbb{E}[|X-X'|^2] \geq \inf_{\eta\in B_\nu} W_2^2(\eta,\mu) = W_2^2(\bar\mu,\mu)$ and there is equality in (9). So $X' = \mathbb{E}[\bar X|X] \sim \bar\mu$ and (X,X') is an optimal coupling. Finally, by definition of conditional expectation, there exists a measurable $T:\mathbb{R}^d \to \mathbb{R}^d$ such that $\mathbb{E}[\bar X|X] = T(X)$ almost surely, which proves the claim.

6.2. **Proof of the main result.** Our proof of Theorem 1.2 follows closely the scheme developed by Gangbo [25] in his alternative proof of Brenier Theorem: first we show the dual attainment, and then we obtain the existence of the transport map $\nabla\varphi$ by doing a first variation around the minimizer f°.

As explained in Remark 1.2, we will treat the case of probability measures with compact supports and the general case separately. Only Items (a) and (b) of Theorem 1.2 are concerned. The proof of Item (c) being identical in the compact and the general case it is proved only once on page 24.

6.2.1. *The compact case.* In this section, we assume that μ and ν have compact supports. In the following theorem we restate Theorem 1.2 (a) and (b) in this special case. In comparison with Theorem 1.2 (a), note that in the compact case f° is always finite valued and bounded from below.

Theorem 6.1. *Let μ and ν be probability measures with compact support on $\mathbb{R}^d$.*

(a) There exists some convex function $f^\circ:\mathbb{R}^d\to\mathbb{R}$ bounded from below such that

$$\overline{\mathcal{T}_2}(\nu|\mu) = \int Q_2 f^\circ\,d\mu - \int f^\circ\,d\nu, \tag{10}$$

where, for any function $g:\mathbb{R}^d\to\mathbb{R}$,

$$Q_2 g(x) = \inf_{y\in\mathbb{R}^d}\{g(y)+|y-x|^2\}, \qquad x\in\mathbb{R}^d.$$

If the supports of μ and ν are contained in the closed ball of radius $R>0$ centered at 0 then f° is $4R$-Lipschitz.

(b) Let h and φ be the convex functions defined by

$$h(x) = \frac{f^\circ(x) + |x|^2}{2}, \quad x \in \mathbb{R}^d \qquad \text{and} \qquad \varphi(y) = h^*(y), \quad y \in \mathbb{R}^d.$$

The function φ is $\mathcal{C}^1$-smooth on $\mathbb{R}^d$ and the map $\nabla\varphi$ is 1-Lipschitz on $\mathbb{R}^d$. The projection $\bar{\mu}$ of μ on B_ν is such that

$$\bar{\mu} = \nabla\varphi_\# \mu.$$

Proof of Item (a) of Theorem 6.1. We will assume that μ and ν are supported in a closed ball B of radius $R > 0$ centered at 0. Let α be the convex function on $[0, \infty)$ defined by

$$\alpha(t) = t^2 \text{ if } t \in [0, 2R] \qquad \text{and} \qquad \alpha(t) = 4Rt - 4R^2 \text{ if } t \geq 2R.$$

and let $\theta(u) = \alpha(|u|)$, for all $u \in \mathbb{R}^d$. Since μ and ν are supported in B and we have the duality statement Theorem 1.1 it is easily seen that

$$\overline{\mathcal{T}}_2(\nu|\mu) = \overline{\mathcal{T}}_\theta(\nu|\mu) = \sup_{g \in \mathcal{G}} \left\{ \int Q_\theta g \, d\mu - \int g \, d\nu \right\}, \tag{11}$$

where the supremum runs over the set $\mathcal{G}$ of convex functions g bounded from below.

Step 1 - Preparation. Let P_θ be the operator acting on functions defined as follows

$$P_\theta g(y) = \sup_{x \in \mathbb{R}^d} \{g(x) - \theta(y - x)\}, \qquad y \in \mathbb{R}^d.$$

First, let us show that we can refine the duality formula (11) by modifying the class of functions $\mathcal{G}$ as follows:

$$\overline{\mathcal{T}}_2(\nu|\mu) = \sup_{f \in \mathcal{F}} \left\{ \int Q_\theta f \, d\mu - \int f \, d\nu \right\},$$

where $\mathcal{F}$ is the subset of $\mathcal{G}$ of functions f which are convex, bounded from below, and satisfy both $f(0) = 0$ and $f = P_\theta(Q_\theta f)$. Indeed, let g be an element of $\mathcal{G}$. Then, by definition of Q_θ, it holds

$$Q_\theta g(x) - \theta(y - x) \leq g(y), \qquad \forall x, y \in \mathbb{R}^d$$

and so $g \geq P_\theta(Q_\theta(g))$. Since $Q_\theta g$ is convex (as an infimum convolution of convex functions), the function f defined by

$$f(y) = P_\theta(Q_\theta g)(y) = \sup_{u \in \mathbb{R}^d} \{Q_\theta g(y - u) - \theta(u)\}, \qquad y \in \mathbb{R}^d$$

is also convex as a supremum of convex functions and bounded from below because P_θ and Q_θ preserve this property. Moreover, it is easily seen that $Q_\theta f = Q_\theta g$ and so $P_\theta(Q_\theta f) = f$. Therefore,

$$\int Q_\theta g \, d\mu - \int g \, d\nu \leq \int Q_\theta f \, d\mu - \int f \, d\nu = \int Q_\theta(f - f(0)) \, d\mu - \int (f - f(0)) \, d\nu,$$

which shows that the duality formula can be restricted to functions $f \in \mathcal{F}$.

Step 2 - Dual attainment. Now, let us show that there is some convex function f° satisfying (10). First of all, if $f \in \mathcal{F}$, then f is $4R$-Lipschitz. This comes from the fact that

$$f(y) = \sup_{x \in \mathbb{R}^d} \{Q_\theta f(x) - \theta(y - x)\}, \qquad \forall y \in \mathbb{R}^d$$

which (the function $y \mapsto \theta(y - x)$ being $4R$-Lipschitz on $\mathbb{R}^d$ for every $x \in \mathbb{R}^d$) shows that f is a supremum of $4R$-Lipschitz functions and is thus $4R$-Lipschitz itself. Since $f(0) = 0$, this implies in particular that $|f| \leq 20R^2$ on $B' = 5B$ (the ball of radius $5R$ centered at 0). Also, since f is $4R$-Lipschitz, it holds

$$f(y) + |y - x|^2 \geq f(x) - 4R|y - x| + |y - x|^2 \qquad \forall x, y \in \mathbb{R}^d.$$

Therefore, $f(y)+|y-x|^2 \geq f(x)$ whenever $|y-x| \geq 4R$. Since $Q_2 f(x) \leq f(x)$, this implies that

$$Q_2 f(x) = \inf_{|y-x|\leq 4R} \{f(y) + |y-x|^2\}, \qquad \forall x \in \mathbb{R}^d.$$

Now, let $f_n \in \mathcal{F}$ be some minimizing sequence. The functions f_n are $4R$-Lipschitz and uniformly bounded on the ball B'. Therefore, it follows from Ascoli's theorem that f_n admits a sub-sequence (still denoted f_n in what follows) converging uniformly to some f° on B'. This function f° is convex on B' as a pointwise limit of convex functions. It is easily seen that $Q_2 f_n \to Q_2 f^\circ$ (uniformly) on B. Since $Q_2 f_n \geq Q_\theta f_n$ pointwise, it holds

$$\overline{\mathcal{T}_2}(\nu|\mu) \geq \int Q_2 f_n \, d\mu - \int f_n \, d\nu \geq \int Q_\theta f_n \, d\mu - \int f_n \, d\nu.$$

Since the right hand side goes to $\overline{\mathcal{T}_2}(\nu|\mu)$ and μ, ν are supported in B, letting $n \to \infty$ yields to (10). For the moment the convex function f° is defined only on B' but it can be easily extended outside B' as follows: the function $\tilde{f}(x) = \inf_{y\in B'}\{f^\circ(y) + 4R|x-y|\}$, $x \in \mathbb{R}^d$, is convex as an infimum convolution of two convex functions, and since f° is $4R$-Lipschitz it is easily seen that $\tilde{f} = f^\circ$ on B'. We can thus assume that f° is a finite valued convex and bounded from below function defined on the whole $\mathbb{R}^d$. This completes the proof. □

In order to prove Item (b) of Theorem 6.1, we will need a technical lemma adapted from [25, Lemma 2.4].

Lemma 6.1. *Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a lower semi-continuous convex function and let $h(x) = \frac{f(x)+|x|^2}{2}$, $x \in \mathbb{R}^d$.*

(a) The function h^* is $\mathcal{C}^1$ smooth and it holds

$$Q_2 f(x) = f(\nabla h^*(x)) + |\nabla h^*(x) - x|^2, \qquad \forall x \in \mathbb{R}^d. \tag{12}$$

(b) For all continuous function $u : \mathbb{R}^d \to \mathbb{R}$ such that $u + f$ is convex and such that $u(x) \geq -a|x| - b$, $x \in \mathbb{R}^d$, for some $a, b \geq 0$ it holds

$$\lim_{t\to 0^+} \frac{Q_2(f+tu)(x) - Q_2(f)(x)}{t} = u(\nabla h^*)(x), \qquad \forall x \in \mathbb{R}^d. \tag{13}$$

Furthermore, there exist some $\alpha, \beta \geq 0$ such that

$$\frac{Q_2(f+tu)(x) - Q_2(f)(x)}{t} \geq -\alpha|x| - \beta, \qquad \forall x \in \mathbb{R}^d, \qquad \forall t \in]0,1].$$

These conclusions also hold for $u_k = -\min(f, k)$, $k \in \mathbb{N}$ in place of u (even though u_k is not necessarily continuous).

Let us admit the lemma until page 21 where the proof is postponed and first prove Item (b) of Theorem 6.1.

Proof of Item (b) of Theorem 6.1. According to Item (a) of Lemma 6.1, the function $\varphi = h^*$ is of class $\mathcal{C}^1$. The fact that $\nabla\varphi$ is 1-Lipschitz follows from Lemma 2.1. Set $\tilde{\mu} = \nabla\varphi_{\#}\mu$ and let us show that $\tilde{\mu} = \bar{\mu}$. First let us prove that $\tilde{\mu} \preceq_c \nu$. Let $u : \mathbb{R}^d \to \mathbb{R}$ be some arbitrary convex function. By optimality of f° it holds, for all $t > 0$,

$$\int Q_2(f^\circ + tu) \, d\mu - \int (f^\circ + tu) \, d\nu \leq \int Q_2(f^\circ) \, d\mu - \int f^\circ \, d\nu$$

Therefore, for all $t > 0$,

$$\int \frac{Q_2(f^\circ + tu) - Q_2(f^\circ)}{t} \, d\mu \leq \int u \, d\nu.$$

Since u is bounded from below by some affine function it satisfies the assumption of Item (b) of Lemma 6.1 and one concludes using Fatou's Lemma that

$$\int u(\nabla\varphi)\,d\mu \le \int u\,d\nu$$

for all convex function $u:\mathbb{R}^d\to\mathbb{R}$. This shows $\tilde\mu \preceq_c \nu$. In particular, one gets $\int f^\circ\,d\tilde\mu \le \int f^\circ\,d\nu$. Actually, for this special function, equality holds. Indeed, the function $u=-f^\circ$ is Lipschitz according to Item (a) of Theorem 6.1 and such that $f^\circ+tu$ is convex for all $0\le t\le 1$. So it satisfies the assumptions of Item (b) of Lemma 6.1. Reasoning as above gives us $\int f^\circ\,d\tilde\mu \ge \int f^\circ\,d\nu$, which shows equality. Therefore,

$$\begin{aligned}\overline{\mathcal{T}}_2(\nu|\mu) &= \int Q_2 f^\circ\,d\mu - \int f^\circ\,d\nu\\ &= \int f^\circ(\nabla\varphi(x)) + |\nabla\varphi(x)-x|^2\,d\mu(x) - \int f^\circ(y)\,d\nu(y)\\ &= \int |\nabla\varphi(x)-x|^2\,d\mu(x)\\ &\ge W_2^2(\tilde\mu,\mu).\end{aligned}$$

Finally, if $\eta\preceq_c\nu$, then

$$\begin{aligned}\overline{\mathcal{T}}_2(\nu|\mu) &= \sup_{g\text{ convex}}\left\{\int Q_2 g\,d\mu - \int g\,d\nu\right\}\\ &\le \sup_{g\text{ convex}}\left\{\int Q_2 g\,d\mu - \int g\,d\eta\right\}\\ &\le W_2^2(\mu,\eta),\end{aligned}$$

where the last inequality follows easily from the inequality $Q_2g(x)-g(y)\le|y-x|^2$, $x,y\in\mathbb{R}^d$. In particular taking $\eta=\tilde\mu$ shows that

$$\overline{\mathcal{T}}_2(\nu|\mu) = W_2^2(\tilde\mu,\mu) = \inf_{\eta\preceq_c\nu} W_2^2(\eta,\mu)$$

and completes the proof. □

Now let us prove Lemma 6.1.

Proof of Lemma 6.1. Let $u:\mathbb{R}^d\to\mathbb{R}$ be a function such that $u+f$ is convex and lower semi-continuous and $u(x)\ge -a|x|-b$, $x\in\mathbb{R}^d$, for some $a,b\ge 0$. Note that for the moment u is not necessarily continuous.
(a) For all $t\in[0,1]$, let h_t be the lower semi-continuous function defined by

$$h_t(x) = \frac{1}{2}\left((1-t)f(x) + t(u(x)+f(x)) + |x|^2\right), \qquad x\in\mathbb{R}^d,$$

with thus $h_0=h$. Applying Lemma 2.1 to $g=h_t$ gives that h_t^* is of class $\mathcal{C}^1$ on $\mathbb{R}^d$ and ∇h_t^* is 1-Lipschitz. Moreover,

$$\begin{aligned}Q_2(f+tu)(x) &= \inf_{y\in\mathbb{R}^d}\{f(y)+tu(y)+|x-y|^2\}\\ &= |x|^2 - 2\sup_{y\in\mathbb{R}^d}\{x\cdot y - h_t(y)\} = |x|^2 - 2h_t^*(x)\end{aligned}$$

and one sees that y is optimal if and only if $x\cdot y = h_t(y)+h_t^*(x)$, that is to say if and only if $y\in\partial h_t^*(x)=\{\nabla h_t^*(x)\}$ (see *e.g* [36, Corollary E 1.4.4]). So it holds, for all $t\in[0,1]$

$$Q_2(f+tu)(x) = f(\nabla h_t^*(x)) + tu(\nabla h_t^*(x)) + |x-\nabla h_t^*(x)|^2, \qquad \forall x\in\mathbb{R}^d, \tag{14}$$

which gives in particular (12) for $t=0$.

(b) Observe that, according to (12), it holds

$$\begin{aligned} Q_2(f+tu)(x) &= \inf_{y\in\mathbb{R}^d}\{(f+tu)(y)+|x-y|^2\} \\ &\le (f+tu)(\nabla h^*(x))+|x-\nabla h^*(x)|^2 \\ &= Q_2 f(x)+tu(\nabla h^*(x)) \end{aligned}$$

and so

$$\frac{Q_2(f+tu)(x)-Q_2f(x)}{t} \le u(\nabla h^*(x)), \qquad \forall x\in\mathbb{R}^d, \forall t\in(0,1]. \tag{15}$$

Similarly, using (14), one easily gets

$$\frac{Q_2(f+tu)(x)-Q_2f(x)}{t} \ge u(\nabla h_t^*(x)), \qquad \forall x\in\mathbb{R}^d, \forall t\in(0,1]. \tag{16}$$

Let us show that for any fixed $x\in\mathbb{R}^d$, $\nabla h_t^*(x)\to\nabla h^*(x)$ as $t\to 0^+$. The function f being convex it is bounded from below by some affine function. Using the lower bound on u, it is easily seen that there exist some constants $c, c'\ge 0$, such that

$$f(y)+tu(y)\ge -c|y|-c', \qquad \forall y\in\mathbb{R}^d.$$

Therefore, since $Q_2(f+tu)(x)\le f(x)+tu(x)\le |f|(x)+|u|(x)$, Identity (14) gives that

$$|f|(x)+|u|(x)\ge -c|\nabla h_t^*|(x)-c'+|x-\nabla h_t^*(x)|^2.$$

Therefore, $\sup_{t\in[0,1]}|\nabla h_t^*|(x)<+\infty$. It follows that if $t_n\in(0,1]$ is some sequence converging to 0, the sequence $\nabla h_{t_n}^*(x)$ admits a subsequence (still denoted $\nabla h_{t_n}^*(x)$ for simplicity) converging to some $\ell(x)$. According to (15) and (16) and the lower bound on u, $Q_2(f+t_nu)(x)\to Q_2f(x)$. On the other hand,

$$Q_2(f+t_nu)(x)=f(\nabla h_{t_n}^*(x))+|x-\nabla h_{t_n}^*(x)|^2+t_nu(\nabla h_{t_n}^*(x)) \tag{17}$$

and so taking the lim inf, it is easily seen (using the lower bound on u) that

$$\liminf_{n\to+\infty} Q_2(f+t_nu)(x)\ge \liminf_{n\to+\infty} f(\nabla h_{t_n}^*(x))+|x-\ell(x)|^2\ge f(\ell(x))+|x-\ell(x)|^2.$$

So we get $Q_2f(x)\ge f(\ell(x))+|x-\ell(x)|^2$ and thus $Q_2f(x)=f(\ell(x))+|x-\ell(x)|^2$ which implies according to the proof of Item (a) that $\ell(x)=\nabla h^*(x)$. It follows that $\nabla h^*(x)$ is the unique limit point of the family $(\nabla h_t^*(x))_{t\in(0,1]}$ when $t\to 0^+$, and so $\nabla h_t^*(x)\to\nabla h^*(x)$ as $t\to 0^+$. To prove (13), let us discuss the different cases.

- If u is continuous, then taking the limit in (15) and (16) gives (13).
- If $u=-\min(f;k)$ for some $k\in\mathbb{N}$, then $u+f=\max(0;f-k)$ is clearly convex and lower semi-continuous and $u\ge -k$. Also, since f is lower bounded by some affine function, the function u is bounded from above by a function of the form $x\mapsto e|x|+e'$, for some $e,e'\ge 0$. Therefore, if $t_n\to 0$, then $t_nu(\nabla h_{t_n}^*(x))\to 0$ and so, taking the limit in (17), one easily gets that sees that $f(\nabla h_{t_n}^*(x))\to f(\nabla h^*(x))$ and so, by definition of u, $u(\nabla h_{t_n}^*(x))\to u(\nabla h^*(x))$ as $n\to+\infty$. Taking the limit in (15) and (16) gives (13) as above.

Finally, in both cases, since $x\mapsto\nabla h_t^*(x)$ is 1-Lipschitz, (16) gives that

$$\frac{Q_2(f+tu)(x)-Q_2f(x)}{t}\ge u(\nabla h_t^*(x))\ge -a|\nabla h_t^*(x)|-b\ge -a|x|-a|\nabla h_t^*(0)|-b$$

and since $\sup_{t\in[0,1]}|\nabla h_t^*|(0)<+\infty$ this completes the proof. □

6.2.2. *Proof of Theorem 1.2 in the general case.* We follow the same path as before and first show the existence of a dual optimizer. As explained in Remark 1.2, another proof has been proposed in [5]. It relies on the fact that Theorem 1.2 on measures with finite second moments can be built by starting from Theorem 6.1 on compact measures.

Proof of Item (a) of Theorem 1.2. First let us justify that $\int Q_2 f \, d\mu$ is finite for any lower semi-continuous convex function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ such that f is integrable with respect to ν. Indeed, selecting some y_o such that $f(y_o)$ is finite one first get that $Q_2 f(x) \leq f(y_o) + |y_o - x|^2$, $x \in \mathbb{R}$. On the other hand, the function $Q_2 f$ is convex on $\mathbb{R}^d$ as an infimum convolution of two convex functions, therefore it is bounded from below by some affine function. Since $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, this shows that $Q_2 f$ is integrable with respect to μ.

Without loss of generality we assume that $\int x \, d\nu(x) = 0$. According to [33, Theorem 2.11 (3)],

$$\overline{\mathcal{T}}_2(\nu|\mu) = \sup_f \left\{ \int Q_2 f \, d\mu - \int f \, d\nu \right\},$$

where the infimum runs over the set of convex functions which are Lipschitz and bounded from below. Without loss of generality one can also impose that $f(0) = 0$. Let $(f_n)_{n\in\mathbb{N}}$ be a maximizing sequence and let $g_n = Q_2 f_n$. Note that, since f_n is convex, g_n is also convex as an infimum convolution of two convex functions.

Let us show that for all $x \in \mathbb{R}^d$, the sequence $g_n(x)$ is bounded. First of all, observe that, for all n, it holds (choosing $y = 0$)

$$g_n(x) = Q_2 f_n(x) = \inf_{y\in\mathbb{R}^d} \{f_n(y) + |y - x|^2\} \leq |x|^2, \qquad \forall x \in \mathbb{R}^d, \tag{18}$$

which shows that for all $x \in \mathbb{R}^d$, the sequence $g_n(x)$ is bounded from above. On the other hand, since $|x|^2 - g_n(x) = \sup_{y\in\mathbb{R}^d}\{2x \cdot y - (f_n(y) + |y|^2)\}$, it follows that the function $|x|^2 - g_n(x)$ is convex as a supremum of convex functions. Therefore, denoting by $b_\mu = \int x \, d\mu$, Jensen inequality yields

$$\int |x|^2 - g_n(x) \, d\mu(x) \geq |b_\mu|^2 - g_n(b_\mu)$$

and so

$$g_n(b_\mu) \geq |b_\mu|^2 - \int |x|^2 \, d\mu(x) + \int g_n(x) \, d\mu(x).$$

Now, observe that $\int g_n \, d\mu \geq \int g_n \, d\mu - \int f_n \, d\nu$, since according to Jensen inequality it holds

$$\int f_n \, d\nu \geq f_n \left(\int x \, d\nu \right) = f_n(0) = 0.$$

Since the sequence $\int g_n \, d\mu - \int f_n \, d\nu$ converges to $\overline{\mathcal{T}}_2(\nu|\mu)$ it is bounded from below. This implies that the sequence $g_n(b_\mu)$ is also bounded from below. Now let x be an arbitrary point in $\mathbb{R}^d$ and write $b_\mu = \frac{1}{2}x + \frac{1}{2}(2b_\mu - x)$, then the convexity of g_n and (18) yield

$$g_n(b_\mu) \leq \frac{1}{2} g_n(x) + \frac{1}{2} g_n(2b_\mu - x) \leq \frac{1}{2} g_n(x) + \frac{1}{2} |2b_\mu - x|^2$$

and so the sequence $g_n(x)$ is bounded from below.

The sequence of convex functions g_n thus satisfies the following boundedness properties:

$$-\infty < \inf_{n\in\mathbb{N}} g_n(x) \qquad \text{and} \qquad \sup_{n\in\mathbb{N}} g_n(x) < +\infty, \qquad \forall x \in \mathbb{R}^d.$$

According to [60, Theorem 10.9 (p.90)], it is possible to extract a subsequence from g_n (that we will still denote by g_n) which converges pointwise to a convex function $g : \mathbb{R}^d \to \mathbb{R}$ (the convergence is also uniform on every compact set, but this will not be needed in the sequel).

According to (18), $g_n(x) \leq |x|^2$ and $\int |x|^2 \, d\mu < +\infty$, so on the one hand Fatou's lemma yields

$$\limsup_{n\to+\infty} \int g_n \, d\mu \leq \int \limsup_{n\to+\infty} g_n \, d\mu = \int g \, d\mu < +\infty.$$

On the other hand,

$$f_n(x) \geq g_n(y) - |y-x|^2, \qquad \forall x, y \in \mathbb{R}^d. \tag{19}$$

Since for a fixed y_o, the sequence $g_n(y_o)$ is bounded from below and $\int |y_o - x|^2 \, d\nu(x) < +\infty$, Fatou's lemma provides

$$\liminf_{n\to+\infty} \int f_n \, d\nu \geq \int \liminf_{n\to+\infty} f_n \, d\nu.$$

Moreover, taking the lim inf in (19) and then optimizing over $y \in \mathbb{R}^d$ gives

$$\liminf_{n\to+\infty} f_n(x) \geq P_2 g(x) = \sup_{y\in\mathbb{R}^d} \{g(y) - |y-x|^2\}, \qquad \forall x \in \mathbb{R}^d.$$

Finally,

$$\begin{aligned}
\overline{\mathcal{T}}_2(\nu|\mu) &= \limsup_{n\to+\infty} \left(\int g_n \, d\mu - \int f_n \, d\mu \right) \\
&\leq \limsup_{n\to+\infty} \int g_n \, d\mu - \liminf_{n\to+\infty} \int f_n \, d\mu \\
&\leq \int g \, d\mu - \int P_2 g \, d\nu.
\end{aligned}$$

Let f° be the function $x \mapsto P_2 g(x) = \sup_{u\in\mathbb{R}^d}\{g(x+u) - |u|^2\} \in \mathbb{R} \cup \{+\infty\}$, $x \in \mathbb{R}^d$. This function is convex and lower semi-continuous as a supremum of convex and continuous functions. Since $g(x) \leq f^\circ(y) + |y-x|^2$ for all $y \in \mathbb{R}^d$, optimizing over $y \in \mathbb{R}^d$ yields $g \leq Q_2 f^\circ$. Therefore,

$$\int Q_2 f^\circ \, d\mu - \int f^\circ \, d\nu \geq \overline{\mathcal{T}}_2(\nu|\mu),$$

which completes the proof. □

Proof of Item (b) of Theorem 1.2. The beginning of the proof of Item (b) of Theorem 6.1 can be repeated exactly as before and yields to the conclusion that $\bar{\mu} := \nabla\varphi_{\#}\mu \preceq_c \nu$. This shows in particular that $\int f^\circ \, d\bar{\mu} \leq \int f^\circ \, d\nu$. To show that this is actually an equality, it is no longer possible to take $u = -f^\circ$, since the function f° is not necessarily Lipschitz. Instead, let us take $u_k = -\min(f^\circ; k)$. The function u_k is such that $f^\circ + tu_k$ is convex for all $t \in [0,1]$, so it is admissible to perform first variation in the optimization problem. Applying Lemma 6.1 and reasoning as in the proof of Theorem 6.1, one sees that $\int f^\circ \, d\bar{\mu} \geq \int \min(f^\circ, k) \, d\bar{\mu} \geq \int \min(f^\circ, k) \, d\nu$. Letting k go to $+\infty$ gives the desired equality. The rest of the proof remains unchanged. □

Proof of Item (c) of Theorem 1.2. Since the probability $\bar{\mu}$ given by Proposition 1.1 is dominated by ν for the convex order, Strassen Theorem implies that there exists a transport kernel q such that $\int y \, dq_x(y) = x$ for $\bar{\mu}$ almost every x and $\nu(\,\cdot\,) = \int q_x(\,\cdot\,) \, d\bar{\mu}(x)$. Let $(X, \bar{X}, Y)$ be a time inhomogeneous Markov chain with initial distribution μ and $\mathrm{Law}(\bar{X}|X) = \delta_{\nabla\varphi(X)}$ and $\mathrm{Law}(Y|\bar{X}) = q_{\bar{X}}$ almost surely, where $\nabla\varphi$ is the transport map given in Theorem 1.2. Then it holds

$$\begin{aligned}
\mathbb{E}[|\mathbb{E}[Y|X] - X|^2] &= \mathbb{E}\left[\left| \mathbb{E}\left[\int y \, dq_{\bar{X}}(y) | X \right] - X \right|^2 \right] = \mathbb{E}\left[\left| \int y \, dq_{\nabla\varphi(X)}(y) - X \right|^2 \right] \\
&= \mathbb{E}\left[|\nabla\varphi(X) - X|^2 \right] = W_2^2(\bar{\mu}, \mu) = \overline{\mathcal{T}}_2(\nu|\mu),
\end{aligned}$$

where the last two equalities come respectively from (5) and Proposition 1.1. This shows the optimality of (X, Y).

Now let $d\pi(x,y) = d\mu(x)dp_x(y)$ be the law of the coupling (X,Y) constructed above (with therefore $dp_x(y) = dq_{\nabla\varphi(x)}(y)$) let $d\pi'(x,y) = d\mu(x)dp'_x(y)$ be another weak optimal transport plan, then

$$\int \left| \int y\, d\left(\frac{p_x + p'_x}{2}\right)(y) - x \right|^2 d\mu(x) \leq \int \frac{|\int y\, dp_x(y) - x|^2 + |\int y\, dp'_x(y) - x|^2}{2}\, d\mu(x) = \overline{\mathcal{T}}_2(\nu|\mu).$$

By optimality and strict convexity of $|\cdot|^2$ we deduce that $\int y\, dp_x = \int y\, dp'_x$ for μ-almost every x. In other words, $\mathbb{E}[Y'|X' = u] = \mathbb{E}[Y|X = u] = \nabla\varphi(u)$ for μ almost every $u \in \mathbb{R}^d$. So if (X', Y') is a weak optimal coupling, one has $\mathbb{E}[Y'|X'] = \nabla\varphi(X')$ almost surely. In particular $\mathbb{E}[Y'|X'] \sim \bar{\mu}$. The fact that $(\mathbb{E}[Y'|X'], Y')$ is a martingale is always true. □

6.3. **Proof of the technical Lemma 2.1.** For the sake of completeness, we recall the proof of this classical duality result for strongly convex functions.

Proof of Lemma 2.1. Let us show that (a) implies (b). First let us show that g^* is finite valued. If x_o is some point where $g(x_o) < +\infty$, then it holds $g^*(y) = \sup_{x\in\mathbb{R}^d}\{x\cdot y - g(x)\} \geq x_o \cdot y - g(x_o)$. This shows that g^* does not take the value $-\infty$. On the other hand, the function $f(x) = g(x) - \frac{|x|^2}{2}$, $x \in \mathbb{R}^d$, is convex by assumption. So there exists some $a \in \mathbb{R}^d$ and $b \in \mathbb{R}$ such that $f(x) \geq a \cdot x + b$ for all $x \in \mathbb{R}^d$. It follows from this that $g(x) = f(x) + \frac{|x|^2}{2} \geq a \cdot x + b + \frac{|x|^2}{2}$, $x \in \mathbb{R}^d$, which easily implies that $g^*(y) < +\infty$ for all $y \in \mathbb{R}^d$. Since $g = f + \frac{|\cdot|^2}{2}$ with f convex, it follows from *e.g.* [36, Theorem E.2.3.1] that the convex conjugate of g is given by

$$g^*(y) = \inf_{x\in\mathbb{R}^d}\left\{f^*(x) + \frac{|y-x|^2}{2}\right\} = \frac{|y|^2}{2} - \sup_{x\in\mathbb{R}^d}\left\{x\cdot y - \left(f^*(x) + \frac{|x|^2}{2}\right)\right\}$$

The function defined by the supremum being clearly convex, it follows that $y \mapsto \frac{|y|^2}{2} - g^*(y)$ is convex on $\mathbb{R}^d$, which shows (b). Conversely, let us show that $(b) \Rightarrow (a)$. Let $k(y) = \frac{|y|^2}{2} - g^*(y)$, $y \in \mathbb{R}^d$, which is convex by assumption. By Fenchel-Legendre duality (see *e.g* [36, Corollary E 1.3.6]), it holds

$$g(x) = \sup_{y\in\mathbb{R}^d}\{x\cdot y - g^*(y)\} = \sup_{y\in\mathbb{R}^d}\left\{x\cdot y - \frac{|y|^2}{2} + k(y)\right\}, \qquad \forall x \in \mathbb{R}^d$$

and so

$$g(x) - \frac{|x|^2}{2} = \sup_{y\in\mathbb{R}^d}\left\{k(y) - \frac{|x-y|^2}{2}\right\} = \sup_{u\in\mathbb{R}^d}\left\{k(x-u) - \frac{|u|^2}{2}\right\} \qquad \forall x \in \mathbb{R}^d.$$

The function $x \mapsto g(x) - \frac{|x|^2}{2}$ is therefore convex as a supremum of convex functions.

Now let us show that (a) implies (c). We have already seen above that g^* is finite valued over $\mathbb{R}^d$. Since $g = f + \frac{|\cdot|^2}{2}$ with f convex, the function g is also strictly convex. Therefore, according to *e.g* [36, Theorem E 4.1.1], it follows that g^* is of class $\mathcal{C}^1$ on $\mathbb{R}^d$.

It remains to prove that ∇g^* is 1-Lipschitz. Since the function $x \mapsto g(x) - \frac{|x|^2}{2}$ is convex, its subgradient is a monotone operator, which means that

$$(b-a)\cdot(y-x) \geq |y-x|^2, \qquad \forall x, y \in \mathbb{R}^d, \qquad \forall b \in \partial g(y), \qquad \forall a \in \partial g(x).$$

Since $u \in \partial g(v)$ is equivalent to $v \in \partial g^*(u) = \{\nabla g^*(u)\}$ (see *e.g.* [36, Corollary E 1.4.4] and [36, Corollary D 2.1.4]), the statement above is equivalent to

$$(\nabla g^*(b) - \nabla g^*(a))\cdot(b-a) \geq |\nabla g^*(b) - \nabla g^*(a)|^2, \qquad \forall a, b \in \mathbb{R}^d,$$

which immediately implies that ∇g^* is 1-Lipschitz. Finally, let us show that (c) implies (b). Since ∇g^* is 1-Lipschitz, it holds

$$(\nabla g^*(y) - \nabla g^*(x))\cdot(y-x) \leq |y-x|^2, \qquad \forall x, y \in \mathbb{R}^d,$$

which easily implies that $x \mapsto \frac{|x|^2}{2} - g^*(x)$ is convex. This completes the proof. □

References

[1] A. Alfonsi, J. Corbetta, and B. Jourdain. Sampling of probability measures in the convex order and approximation of Martingale Optimal Transport problems. *ArXiv e-prints*, page arXiv:1709.05287, Sept. 2017. 4, 15

[2] J.-J. Alibert, G. Bouchitté, and T. Champion. A new class of cost for optimal transport planning. Preprint, Mar. 2018. 6, 7, 16

[3] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008. 2, 17

[4] L. Ambrosio and S. Rigot. Optimal mass transportation in the Heisenberg group. *J. Funct. Anal.*, 208(2):261–301, 2004. 2

[5] J. Backhoff Veraguas, M. Beiglboeck, and G. Pammer. Existence and Cyclical monotonicity for weak transport costs. *ArXiv e-prints*, Sept. 2018. 5, 9, 23

[6] D. Bakry, I. Gentil, and M. Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014. 2

[7] M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *Ann. Probab.*, 44(1):42–106, 2016. 6, 11, 12

[8] M. Beiglboeck and N. Juillet. Shadow couplings. *ArXiv e-prints*, page arXiv:1609.03340, Sept. 2016. 2, 16

[9] J. Bertrand. Existence and uniqueness of optimal maps on Alexandrov spaces. *Adv. Math.*, 219(3):838–851, 2008. 2

[10] C. Boubel and N. Juillet. The Markov-quantile process attached to a family of Marginals. *ArXiv e-prints*, page arXiv:1804.10514, Apr. 2018. 2

[11] M. Bowles and N. Ghoussoub. A Theory of Transfers: Duality and convolution. *ArXiv e-prints*, Apr. 2018. 7

[12] Y. Brenier. Décomposition polaire et réarrangement monotone des champs de vecteurs. *C. R. Acad. Sci. Paris Sér. I Math.*, 305(19):805–808, 1987. 1

[13] Y. Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.*, 44(4):375–417, 1991. 1

[14] L. A. Caffarelli. Monotonicity properties of optimal transportation and the FKG and related inequalities. *Comm. Math. Phys.*, 214(3):547–563, 2000. 1, 5, 7

[15] G. Conforti. A second order equation for Schrödinger bridges with applications to the hot gas experiment and entropic transportation cost. *ArXiv e-prints*, Apr. 2017. 6

[16] G. Conforti and L. Ripani. Around the entropic Talagrand inequality. *ArXiv e-prints*, page arXiv:1809.02062, Sept. 2018. 6

[17] D. Constantinos, D. Alan, and T. Christos. Strong duality for a multiple-good monopolist. *Econometrica*, 85(3):735–767, 2017. 16

[18] D. Cordero-Erausquin. Some applications of mass transport to Gaussian-type inequalities. *Arch. Ration. Mech. Anal.*, 161(3):257–269, 2002. 7

[19] H. De March. Local structure of multi-dimensional martingale optimal transport. *ArXiv e-prints*, page arXiv:1805.09469, May 2018. 6

[20] A. Dembo. Information inequalities and concentration of measure. *Ann. Probab.*, 25(2):927–939, 1997. 6

[21] S. Fang and J. Shao. Optimal transport maps for Monge-Kantorovich problem on loop groups. *J. Funct. Anal.*, 248(1):225–257, 2007. 2

[22] M. Fathi and Y. Shu. Curvature and transport inequalities for Markov chains in discrete spaces. *Bernoulli*, 24(1):672–698, 2018. 6

[23] D. Feyel and A. S. Üstünel. Monge-Kantorovitch measure transportation and Monge-Ampère equation on Wiener space. *Probab. Theory Related Fields*, 128(3):347–385, 2004. 2

[24] A. Figalli and L. Rifford. Mass transportation on sub-Riemannian manifolds. *Geom. Funct. Anal.*, 20(1):124–159, 2010. 2

[25] W. Gangbo. An elementary proof of the polar factorization of vector-valued functions. *Arch. Rational Mech. Anal.*, 128(4):381–399, 1994. 18, 20

[26] W. Gangbo and R. J. McCann. The geometry of optimal transportation. *Acta Math.*, 177(2):113–161, 1996. 2, 8

[27] N. Ghoussoub, Y.-H. Kim, and T. Lim. Structure of optimal martingale transport plans in general dimensions. *ArXiv e-prints*, page arXiv:1508.01806, Aug. 2015. 6

[28] N. Gigli. On the inverse implication of Brenier-McCann theorems and the structure of $(\mathcal{P}_2(M), W_2)$. *Methods Appl. Anal.*, 18(2):127–158, 2011. 2, 4

[29] N. Gigli, T. Rajala, and K.-T. Sturm. Optimal maps and exponentiation on finite-dimensional spaces with Ricci curvature bounded from below. *J. Geom. Anal.*, 26(4):2914–2929, 2016. 2

[30] N. Gozlan and C. Léonard. Transport inequalities. A survey. *Markov Process. Related Fields*, 16(4):635–736, 2010. 6

[31] 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(3):1667–1693, 2018. 4, 6, 15

[32] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Displacement convexity of entropy and related inequalities on graphs. *Probab. Theory Related Fields*, 160(1-2):47–94, 2014. 6

[33] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Kantorovich duality for general transport costs and applications. *J. Funct. Anal.*, 273(11):3327–3405, 2017. 1, 2, 3, 6, 17, 23

[34] G. Hargé. Inequalities for the Gaussian measure and an application to Wiener space. *C. R. Acad. Sci. Paris Sér. I Math.*, 333(8):791–794, 2001. 7

[35] P. Henry-Labordère and N. Touzi. An explicit martingale version of the one-dimensional Brenier theorem. *Finance Stoch.*, 20(3):635–668, 2016. 6

[36] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of convex analysis*. Grundlehren Text Editions. Springer-Verlag, Berlin, 2001. 7, 21, 25

[37] F. Hirsch, C. Profeta, B. Roynette, and M. Yor. *Peacocks and associated martingales, with explicit constructions*, volume 3 of *Bocconi & Springer Series*. Springer, Milan; Bocconi University Press, Milan, 2011. 2

[38] F. Hirsch and B. Roynette. On $\mathbb{R}^d$-valued peacocks. *ESAIM Probab. Stat.*, 17:444–454, 2013. 2

[39] F. Hirsch, B. Roynette, and M. Yor. Kellerer's theorem revisited. In Springer, editor, *Asymptotic Laws and Methods in Stochastics. Volume in Honour of Miklos Csorgo*, Fields Institute Communications Series, 2014. 2

[40] F. Hirsch and M. Yor. Looking for martingales associated to a self-decomposable law. *Electron. J. Probab.*, 15:no. 29, 932–961, 2010. 2

[41] D. Hobson and M. Klimmek. Robust price bounds for the forward starting straddle. *Finance Stoch.*, 19(1):189–214, 2015. 6

[42] D. Hobson and A. Neuberger. Robust bounds for forward start options. *Math. Finance*, 22(1):31–56, 2012. 6

[43] D. G. Hobson and D. Norgilas. The left-curtain martingale coupling in the presence of atoms. *to appear in Ann. of Appl. Probab.*, 2018. 6

[44] N. Juillet. Peacocks parametrised by a partially ordered set. In *Séminaire de Probabilités XLVIII*, volume 2168 of *Lecture Notes in Math.*, pages 13–32. Springer, Cham, 2016. 2

[45] N. Juillet. Stability of the shadow projection and the left-curtain coupling. *Ann. Inst. Henri Poincaré Probab. Stat.*, 52(4):1823–1843, 2016. 6

[46] H. G. Kellerer. Markov-Komposition und eine Anwendung auf Martingale. *Math. Ann.*, 198:99–122, 1972. 2

[47] H. G. Kellerer. Integraldarstellung von Dilationen. In *Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes (Tech. Univ., Prague, 1971; dedicated to the memory of Antonín Špaček)*, pages 341–374. Academia, Prague, 1973. 2

[48] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001. 6

[49] C. Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, 34(4):1533–1574, 2014. 6

[50] T. Lim. Optimal martingale transport between radially symmetric marginals in general dimensions. *ArXiv e-prints*, page arXiv:1412.3530, Dec. 2014. 6

[51] G. Lowther. Limits of one-dimensional diffusions. *Ann. Probab.*, 37(1):78–106, 2009. 2

[52] K. Marton. Bounding $\overline{d}$-distance by informational divergence: a method to prove measure concentration. *Ann. Probab.*, 24(2):857–866, 1996. 6

[53] K. Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.*, 6(3):556–571, 1996. 6

[54] R. J. McCann. Polar factorization of maps on Riemannian manifolds. *Geom. Funct. Anal.*, 11(3):589–608, 2001. 2

[55] R. J. McCann and N. Guillen. Five lectures on optimal transportation: geometry, regularity and applications. In *Analysis and geometry of metric measure spaces*, volume 56 of *CRM Proc. Lecture Notes*, pages 145–180. Amer. Math. Soc., Providence, RI, 2013. 1

[56] E. Milman. Spectral estimates, contractions and hypercontractivity. *J. Spectr. Theory*, 8(2):669–714, 2018. 7

[57] M. Nutz and F. Stebegg. Canonical supermartingale couplings. *Ann. Probab.*, 46(6):3351–3398, 2018. 6, 12

[58] E. Outerelo and J. M. Ruiz. *Mapping degree theory*, volume 108 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI; Real Sociedad Matemática Española, Madrid, 2009. 14

[59] S. T. Rachev and L. Rüschendorf. *Mass transportation problems. Vol. I.* Probability and its Applications (New York). Springer-Verlag, New York, 1998. Theory. 8

[60] R. T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970. 23

[61] L. Rüschendorf and S. T. Rachev. A characterization of random variables with minimum L^2-distance. *J. Multivariate Anal.*, 32(1):48–54, 1990. 8

[62] P.-M. Samson. Concentration of measure inequalities for Markov chains and Φ-mixing processes. *Ann. Probab.*, 28(1):416–461, 2000. 6

[63] P.-M. Samson. Concentration inequalities for convex functions on product spaces. In *Stochastic inequalities and applications*, volume 56 of *Progr. Probab.*, pages 33–52. Birkhäuser, Basel, 2003. 6

[64] P.-M. Samson. Infimum-convolution description of concentration properties of product probability measures, with applications. *Ann. Inst. H. Poincaré Probab. Statist.*, 43(3):321–338, 2007. 6

[65] P.-M. Samson. Concentration of measure principle and entropy inequalities. In *Convexity and concentration*, volume 161 of *The IMA Volumes in Mathematics and its Applications*, pages 55–105. Springer, New York, 2017. 4, 6

[66] P.-M. Samson. Transport-entropy inequalities on locally acting groups of permutations. *Electron. J. Probab.*, 22:Paper No. 62, 33, 2017. 6

[67] Y. Shu. From Hopf-Lax formula to optimal weak transfer plan. *ArXiv e-prints*, page arXiv:1609.03405, Sept. 2016. 7

[68] Y. Shu. Hamilton-Jacobi equations on graph and applications. *Potential Anal.*, 48(2):125–157, 2018. 6

[69] V. Strassen. The existence of probability measures with given marginals. *Ann. Math. Statist.*, 36:423–439, 1965. 2, 10

[70] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Inst. Hautes Études Sci. Publ. Math.*, (81):73–205, 1995. 6

[71] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996. 6

[72] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003. 2, 8, 17

[73] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new. 2, 3, 8

NG : Université Paris Descartes, MAP5, UMR 8145, 45 rue des Saints Pères, 75270 Paris Cedex 06

E-mail address: natael.gozlan@parisdescartes.fr

NJ: Université de Strasbourg, IRMA, UMR 7501, 7 rue René-Descartes 67084 Strasbourg Cedex

E-mail address: nicolas.juillet@math.unistra.fr