Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2018, Vol. 54, No. 3, 1667–1693
https://doi.org/10.1214/17-AIHP851

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Characterization of a class of weak transport-entropy inequalities on the line[1]

Nathael Gozlan[a], Cyril Roberto[b], Paul-Marie Samson[a], Yan Shu[b] and Prasad Tetali[c,2]

[a]*Laboratoire d'Analyse et de Mathématiques Appliquées (UMR CNRS 8050), Université Paris Est Marne la Vallée, 5 bd Descartes, 77454 Marne la Vallée Cedex 2, France. E-mail: natael.gozlan@u-pem.fr; paul-marie.samson@u-pem.fr*
[b]*Laboratoire Modal'X, EA 3454, Université Paris Ouest Nanterre La Défense, 200 avenue de la République, 92000 Nanterre, France. E-mail: c.roberto@math.cnrs.fr; yan.shu@u-paris10.fr*
[c]*School of Mathematics & School of Computer Science, Georgia Institute of Technology, Atlanta, GA 30332, USA. E-mail: tetali@math.gatech.edu*

Received 13 January 2016; revised 6 June 2017; accepted 4 July 2017

Abstract. We study an weak transport cost related to the notion of convex order between probability measures. On the real line, we show that this weak transport cost is reached for a coupling that does not depend on the underlying cost function. As an application, we give a necessary and sufficient condition for weak transport-entropy inequalities (related to concentration of convex/concave functions) to hold on the line. In particular, we obtain a weak transport-entropy form of the convex Poincaré inequality in dimension one.

Résumé. Dans cet article, nous étudions une nouvelle famille de coûts de transport optimaux faibles en lien avec la notion d'ordre convexe pour les mesures de probabilité. Nous montrons, en dimension un, que le couplage optimal ne dépend pas de la fonction de coût choisie. Nous utilisons ensuite ce résultat pour établir une condition nécessaire et suffisante pour les inégalités de transport-entropie associées à ces coûts de transport faibles. En particulier, nous obtenons une forme transport équivalente de l'inégalité de Poincaré restreinte aux fonctions convexes sur la droite.

MSC: 60E15; 32F32; 26D10

Keywords: Transport inequalities; Concentration of measure; Majorization

1. Introduction

The aim of this paper is to study a weak transport cost and its associated weak transport-entropy inequality, both introduced by four of the authors in [25], on the real line. In order to present our results, we shall first introduce the various mathematical objects of interest to us, placing and motivating their significance within the classical theory of optimal transport and its connection with the concentration of measure phenomenon.

Throughout the paper, $\mathcal{P}(\mathbb{R})$ denotes the set of Borel probability measures on $\mathbb{R}$ and $\mathcal{P}_1(\mathbb{R}) := \{\mu \in \mathcal{P}(\mathbb{R}) : \int_{\mathbb{R}} |x|\mu(dx) < \infty\}$, the subset of probability measures having a finite first moment.

[1]Supported by the Grants ANR 2011 BS01 007 01, ANR 10 LABX-58 and ANR11-LBX-0023-01.
[2]Supported by the NSF Grants DMS 1101447 and 1407657.

Let $\theta : \mathbb{R}^+ \to \mathbb{R}^+$, with $\theta(0) = 0$, be a measurable function referred to as the *cost* function. Then, the usual optimal *transport cost*, in the sense of Kantorovich, between two probability measures μ and ν on $\mathbb{R}$ is defined by

$$\mathcal{T}_\theta(\nu, \mu) := \inf_\pi \iint \theta\big(|x - y|\big)\pi(dx\, dy), \tag{1}$$

where the infimum runs over the set of couplings π between μ and ν, i.e., probability measures on $\mathbb{R}^2$ such that $\pi(dx \times \mathbb{R}) = \mu(dx)$ and $\pi(\mathbb{R} \times dy) = \nu(dy)$.

Since the works by Marton [34–36] and Talagrand [47], these transport costs have been extensively used as a tool to reach concentration properties for measures on product spaces. More precisely, optimal transport is related to the concentration of measure phenomenon via the so-called transport-entropy inequalities that we now recall. A probability measure μ on $\mathbb{R}$ is said to satisfy the transport-entropy inequality $\mathrm{T}(\theta)$, if for all $\nu \in \mathcal{P}(\mathbb{R})$, it holds

$$\mathcal{T}_\theta(\nu, \mu) \le \mathrm{H}(\nu|\mu), \tag{2}$$

where $\mathrm{H}(\nu|\mu)$ denotes the relative entropy (also called Kullback–Leibler distance) of ν with respect to μ, defined by

$$\mathrm{H}(\nu|\mu) := \int \log\left(\frac{d\nu}{d\mu}\right) d\nu,$$

if ν is absolutely continuous with respect to μ, and $\mathrm{H}(\nu|\mu) := \infty$ otherwise. Note that we focus on the line, but that all the above definitions easily generalize to probability measures on a general metric space. As a special case, as proved by Talagrand in his seminal paper [47], Inequality (2) is satisfied by the standard Gaussian measure for the cost $\theta(x) = x^2/2$. By extension, we shall say that μ satisfies the inequality $\mathrm{T}_2(C)$ (often referred to as "Talagrand's inequality" in the literature), if (2) holds for a cost function of the form $\theta(x) = x^2/C$, for some $C > 0$. We refer to the books or survey [13,22,31,48] for a complete presentation of transport-entropy inequalities and of the concentration of measure phenomenon, as well as for bibliographic references in the field.

In the next few lines we shall shortly discuss the consequences of transport-entropy inequalities in terms of concentration. For simplicity we may only consider Inequality $\mathrm{T}_2(C)$.

As discovered by Marton and Talagrand, when a probability μ satisfies $\mathrm{T}_2(C)$, then for all positive integers n, and all functions $f : \mathbb{R}^n \to \mathbb{R}$ which are 1-Lipschitz with respect to the Euclidean norm on $\mathbb{R}^n$, it holds

$$\mu^n\big(f > \mathrm{med}(f) + r\big) \le e^{-(r-r_o)^2/C}, \quad \forall r \ge r_o := \sqrt{C\log(2)}, \tag{3}$$

where $\mathrm{med}(f)$ denotes a median of f under μ^n. We refer to [13,31] for a presentation of the numerous applications of this type of dimension-free concentration of measure inequalities. Conversely, it was shown by the first named author in [20] that a probability μ satisfying the dimension-free Gaussian concentration (3) necessarily satisfies $\mathrm{T}_2(C)$, thus giving to Inequality T_2 a special status among other functional inequalities appearing in the concentration of measure literature. The key argument explaining why Talagrand's inequality implies the dimension-free concentration behavior (3) is the well-known tensorisation property enjoyed in general by inequalities of the form $\mathrm{T}(\theta)$ (see, e.g., [22]). The tensorisation property shows in particular that if μ satisfies $\mathrm{T}_2(C)$, then the n-fold product measure $\mu \otimes \cdots \otimes \mu$ also satisfies $\mathrm{T}_2(C)$ (on $\mathbb{R}^n$) with the same constant C.

More generally, given a measure on a product space (which is not necessarily a product measure), and assuming that each of its conditional one-dimensional marginals satisfies a transport-entropy inequality, several authors have obtained, using different non-independent tensorisation strategies, transport-entropy inequalities for the whole measure under weak dependence assumptions (see for instance [16,37,49,50]). Then, the transport-entropy inequality for the whole measure leads again to concentration properties using the same classical arguments as in the product case. Thus, in many situations one is reduced to verifying the one-dimensional transport-entropy inequalities, and therefore it is of a real interest to characterize those probability measures μ on $\mathbb{R}$ that satisfy the inequality T_2, and more generally, $\mathrm{T}(\theta)$ for a general cost function θ.

In this direction, in [21], the first-named author obtained necessary and sufficient conditions for $\mathrm{T}(\theta)$ to hold, when the cost function $\theta : \mathbb{R}^+ \to \mathbb{R}^+$ is continuous, convex and quadratic near 0. In order to present such conditions, we need to introduce some notation. Denote by $F_\mu(x) := \mu((-\infty, x])$, $x \in \mathbb{R}$, the cumulative distribution function of a probability measure μ and by F_μ^{-1} its general inverse defined by

$$F_\mu^{-1}(u) := \inf\big\{x \in \mathbb{R},\, F_\mu(x) \geq u\big\} \in \mathbb{R} \cup \{\pm\infty\}, \quad \forall u \in [0,1].$$

The conditions obtained in [21] are expressed in terms of the behavior of the modulus of continuity of the non-decreasing map $U_\mu := F_\mu^{-1} \circ F_\tau$, where τ is the symmetric exponential distribution on $\mathbb{R}$,

$$\tau(dx) := \frac{1}{2} e^{-|x|}\, dx,$$

so that

$$U_\mu(x) = \begin{cases} F_\mu^{-1}(1 - \frac{1}{2}e^{-|x|}) & \text{if } x \geq 0, \\ F_\mu^{-1}(\frac{1}{2}e^{-|x|}) & \text{if } x \leq 0. \end{cases} \tag{4}$$

By construction, U_μ is the unique left-continuous and non-decreasing map transporting τ onto μ (i.e., $\int f \circ U_\mu \, d\tau = \int f \, d\mu$ for all f). In the special case of the inequality T_2, the characterization of [21] reads as follows: a probability measure μ satisfies $\mathrm{T}_2(C)$ for some constant C if and only if the following holds

- for some constant $b > 0$ and all $u \geq 0$, it holds

$$\sup_{x \in \mathbb{R}} \big(U_\mu(x+u) - U_\mu(x)\big) \leq \frac{1}{b}\sqrt{1+u},$$

- for some constant $c > 0$ and all f of class $\mathcal{C}^1$, the Poincaré inequality holds

$$\mathrm{Var}_\mu(f) \leq c \int f'^2 \, d\mu. \tag{5}$$

We refer to [21] for a precise quantitative relation between C, b and c.

In the present paper, partly following [21], we focus on the study of a new weak transport-entropy inequality introduced in [25] that is related to a weak type of dimension-free concentration. More precisely, in dimension one, we consider the weak optimal transport cost of ν with respect to μ defined by

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

where the infimum runs over all couplings $\pi(dx\,dy) = p(x,dy)\mu(dx)$ of μ and ν, and where $p(x,\cdot)$ denotes the disintegration kernel of π with respect to its first marginal. The notation bar comes from the barycenter entering in its definition. Note that, contrary to the usual transport cost, $\overline{\mathcal{T}}_\theta$ is not symmetric. Also, in terms of random variables, one has the following interpretation $\overline{\mathcal{T}}_\theta(\nu|\mu) = \inf \mathbb{E}(\theta(|X - \mathbb{E}(Y|X)|))$ whereas $\mathcal{T}_\theta(\nu,\mu) = \inf \mathbb{E}(\theta(|X-Y|))$, where in both cases the infimum runs over all random variables X, Y such that X follows the law μ and Y the law ν. As a consequence, when θ is convex, by Jensen's inequality, one has $\overline{\mathcal{T}}_\theta(\nu|\mu) \leq \mathcal{T}_\theta(\nu,\mu)$. Therefore, if a measure μ satisfies $\mathrm{T}(\theta)$ then it also satisfies the following weaker transport-entropy inequalities.

Definition 1.1. Let $\theta : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex cost function. A probability measure μ on $\mathbb{R}$ is said to satisfy the weak transport-entropy inequality

$\overline{\mathrm{T}}^+(\theta)$: if for all $\nu \in \mathcal{P}_1(\mathbb{R})$, it holds

$$\overline{\mathcal{T}}_\theta(\nu|\mu) \leq \mathrm{H}(\nu|\mu),$$

$\overline{\mathrm{T}}^{-}(\theta)$: if for all $\nu \in \mathcal{P}_1(\mathbb{R})$, it holds

$$\overline{\mathcal{T}}_\theta(\mu|\nu) \le \mathrm{H}(\nu|\mu), \quad \text{and}$$

$\overline{\mathrm{T}}(\theta)$: if μ satisfies both $\overline{\mathrm{T}}^{+}(\theta)$ and $\overline{\mathrm{T}}^{-}(\theta)$.

In Section 4, we recall, from [25, Section 4] a dual formulation of these weak transport inequalities in terms of infimum convolution operators. In particular, the inequality $\overline{\mathrm{T}}(\theta)$ appears as the dual formulation of the so-called convex (τ)-property introduced by Maurey [38] (see also [40] for further development).

The above defined weak transport-entropy inequalities are of particular interest since the class of measures satisfying such inequalities also includes discrete measures on $\mathbb{R}$ such as Bernoulli, binomial and Poisson measures [25,40]. In comparison, the classical Talagrand's transport inequality (and more generally any $\mathrm{T}(\theta)$) is never satisfied by a discrete probability measure unless it is a Dirac measure.[3] Moreover these weak transport-entropy inequalities also enjoy the tensorisation property (see [25, Theorem 4.11]), thus connecting them to a special dimension-free concentration behavior. For instance, let us recall the following particular case of [25, Corollary 5.11].

Theorem 1.2. *A probability measure μ satisfies $\overline{\mathrm{T}}_2(C)$ (i.e., $\overline{\mathrm{T}}(\theta)$ with $\theta(x) = x^2/C$, $x \in \mathbb{R}$) if and only if it satisfies the following weak dimension-free concentration of measure inequality: there exist $r_o \ge 0$ and $C' > 0$ such that for all positive integers n and all* convex *or* concave *functions $f : \mathbb{R}^n \to \mathbb{R}$ which are also* 1*-Lipschitz for the Euclidean norm on $\mathbb{R}^n$, it holds*

$$\mu^n\big(f > \mathrm{med}(f) + r\big) \le e^{-(r-r_o)^2/C'}, \quad \forall r \ge r_o, \tag{6}$$

where $\mathrm{med}(f)$ *denotes a median of f under μ.*

We refer to [25, Corollary 5.11] for more details about the quantitative links between C and r_o, C' in both directions.

The first example of a probability measure satisfying this weak dimension-free concentration of measure phenomenon for convex or concave functions (6) is the uniform measure μ on the two point space $\{0, 1\}$ as was discovered by Talagrand in [45]. Soon after, Johnson and Schechtman [30] extended Talagrand's induction proof to a more general setting and showed that (6) was in fact true as soon as μ is a probability measure on a normed space whose support has a finite diameter. This result was then recovered by Maurey [38] using the formalism of convex infimum convolution inequalities. We refer to [31,46] for a survey of subsequent developments and applications of this type of concentration inequalities.

A natural question in this context was to determine the set of all probability measures on the real line satisfying the concentration property (6). Even if some sufficient conditions already appeared in [2,3] using an approach based on the logarithmic Sobolev inequality restricted to convex functions, the question was still open and was the main motivation behind the present paper.

With this motivation in mind we shall prove the following characterization (the main result of the paper) of the transport inequalities $\overline{\mathrm{T}}(\theta)$ associated to any convex cost function θ which is quadratic near 0. In the sequel we may use the following standard notation $\theta(a\cdot)$ for the function $x \mapsto \theta(ax)$.

Theorem 1.3. *Let $\mu \in \mathcal{P}_1(\mathbb{R})$, $t_o > 0$ and $\theta : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex cost function such that $\theta(t) = t^2$ for all $t \le t_o$. The following propositions are equivalent*:

(i) *There exists $a > 0$ such that μ satisfies $\overline{\mathrm{T}}(\theta(a\cdot))$.*
(ii) *There exists $b > 0$ such that for all $u > 0$,*

$$\sup_x \big(U_\mu(x+u) - U_\mu(x)\big) \le \frac{1}{b}\theta^{-1}\big(u + t_o^2\big),$$

where $\theta^{-1} : \mathbb{R}^+ \to \mathbb{R}^+$ denotes the inverse of the (*increasing*) *function θ.*

[3] Indeed, as mentioned above, the Poincaré inequality is a consequence of Talagrand's transport inequality, and as such it forces the support of μ to be connected.

Moreover, constants are related as follows:

(i) *implies* (ii) *with* $b = a\kappa_1$,
(ii) *implies* (i), *with* $a = b\kappa_2$,

where $\kappa_1 := \frac{t_o}{8\theta^{-1}(\log(3)+t_o^2)}$ *and* $\kappa_2 := \frac{\min(1,t_o)}{210\theta^{-1}(2+t_o^2)}$.

In particular, a probability measure μ on $\mathbb{R}$ satisfies the weak dimension-free concentration of measure inequality (6) if and only if there is some $b > 0$ such that

$$\sup_x \big(U_\mu(x+u) - U_\mu(x)\big) \le \frac{1}{b}\sqrt{u+1}, \quad \forall u \ge 0.$$

Remark 1.4. In comparison with the characterization of the inequalities T(θ) given in [21], one sees that only the condition on the modulus of continuity of U_μ remains. Nevertheless, as we shall explain below, the Poincaré Inequality has not completely disappeared from the picture. Also, denoting by Δ_μ the modulus of continuity of U_μ defined by

$$\Delta_\mu(h) = \sup\big\{U_\mu(x+u) - U_\mu(x), x \in \mathbb{R}, 0 \le u \le h\big\}, \quad h \ge 0,$$

condition (ii) asserts that

$$\Delta_\mu(h) \le \frac{1}{b}\theta^{-1}\big(h + t_o^2\big).$$

Therefore Δ_μ is bounded above near zero but does not necessarily go to zero, as h goes to zero. In fact, if the measure μ is discrete and not a Dirac measure, the support of μ is not connected and so there exist $a < b$ with a and b in the support of μ such that $\mu(]a, b[) = 0$. In that case, we may easily check that for all $h > 0$, $b - a \le \Delta_\mu(h)$. This shows that in a discrete setting $\lim_{h\to 0} \Delta_\mu(h) > 0$.

The proof of Theorem 1.3, given in Section 6, is based on a refined study of the weak transport cost $\overline{\mathcal{T}}_\theta(\mu|\nu)$ of independent interest. Indeed, we shall prove that, in dimension 1, all the optimal weak transport costs $\overline{\mathcal{T}}_\theta(\mu|\nu)$ are achieved by the same coupling independently of the convex cost function θ. This result is well-known for the classical transport cost $\mathcal{T}_\theta$. More precisely, it follows from the works by Hoeffding, Fréchet and Dall'Aglio [15,19,29] (see also [14]) that $\mathcal{T}_\theta(\mu,\nu) = \int \theta(|x - T_{\nu,\mu}(x)|)\nu(dx)$ where $T_{\nu,\mu} := F_\mu^{-1} \circ F_\nu$. In particular, given any two convex costs θ_1, θ_2, it holds $\mathcal{T}_{\theta_1+\theta_2}(\mu,\nu) = \mathcal{T}_{\theta_1}(\mu,\nu) + \mathcal{T}_{\theta_2}(\mu,\nu)$. In order to state our second main result, recall that one says that ν_1 is dominated by ν_2 in the convex order, denoted by $\nu_1 \preceq \nu_2$, if $\int f\,d\nu_1 \le \int f\,d\nu_2$ for all convex functions $f : \mathbb{R} \to \mathbb{R}$.

Theorem 1.5. *For all* $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$, *there exists a probability measure* $\hat{\gamma}$ *dominated by* ν *in the convex order,* $\hat{\gamma} \preceq \nu$, *such that* for all *convex cost functions* θ *it holds*

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

In particular, for any two convex cost functions θ_1, θ_2, *it holds*

$$\overline{\mathcal{T}}_{\theta_1+\theta_2}(\mu|\nu) = \overline{\mathcal{T}}_{\theta_1}(\mu|\nu) + \overline{\mathcal{T}}_{\theta_2}(\mu|\nu). \tag{7}$$

The notion of convex ordering, characterized by Strassen [43] in terms of martingales, will turn out to be crucial in the understanding of the weak transport costs $\overline{\mathcal{T}_\theta}$. In Section 2, we recall certain classical properties of the convex order and in particular its geometrical meaning (in discrete setting) given by Rado's theorem [39] (see Theorem 2.9). From this geometrical interpretation, we shall obtain an intermediate outcome (Theorem 2.10) that might be interpreted as a discrete version of Theorem 1.5. Finally, the proof of Theorem 1.5, given in Section 3, will follow by an approximation argument.

With the result of Theorem 1.5 in hand, we can briefly introduce the main ideas of the proof of Theorem 1.3. Following [21], the weak transport-entropy inequality (i) will follow from (ii) by decomposition of the weak optimal

cost θ into two parts. One part is related to the quadratic behavior of θ on $[0, t_o]$ and the second part is related to its behavior for $t \geq t_o$: one has $\theta \leq \theta_1 + \theta_2$ with

$$\theta_1(t) := t^2 \mathbf{1}_{[0,t_o]}(t) + \big(2tt_o - t_o^2\big)\mathbf{1}_{[t_o,+\infty)}(t),$$

and

$$\theta_2(t) := \big[\theta(t) - t_o^2\big]_+ = \big(\theta(t) - t_o^2\big)\mathbf{1}_{[t_o,+\infty)}(t), \quad t \in \mathbb{R}.$$

Therefore, by Theorem 1.5,

$$\overline{\mathcal{T}}_{\theta(a\cdot)}(\nu|\mu) \leq \overline{\mathcal{T}}_{(\theta_1+\theta_2)(a\cdot)}(\nu|\mu) = \overline{\mathcal{T}}_{\theta_1(a\cdot)}(\nu|\mu) + \overline{\mathcal{T}}_{\theta_2(a\cdot)}(\nu|\mu) \leq 2\mathrm{H}(\nu|\mu),$$

for a proper choice of the constant a, where the last inequality will follow by relating the condition appearing in (ii) to the two weak transport-entropy inequalities with cost $\theta_1(a\cdot)$ and $\theta_2(a\cdot)$. More precisely, following [21, Theorem 2.2], we will show in Theorem 6.1 that (ii) characterizes the weak transport-entropy inequality $\overline{\mathrm{T}}(\theta_2(a\cdot))$ while, as stated in the next theorem (our last main result), the weaker condition $\sup_x(U_\mu(x+1) - U_\mu(x)) \leq h$, for some $h > 0$, characterizes the weak transport-entropy inequality $\overline{\mathrm{T}}(\theta_1(a\cdot))$ which thus appears (thank to [8]) to be also equivalent to the Poincaré inequality (5) restricted to convex functions.

Theorem 1.6. *Let $\mu \in \mathcal{P}_1(\mathbb{R})$. The following assertions are equivalent*:

(i) *There exists $h > 0$ such that*

$$\sup_{x\in\mathbb{R}}\big[U_\mu(x+1) - U_\mu(x)\big] \leq h.$$

(ii) *There exists $C > 0$ such that for all convex function f on $\mathbb{R}$ it holds*

$$\mathrm{Var}_\mu(f) \leq C \int_{\mathbb{R}} |\nabla f|^2 \, d\mu,$$

where, for all $x \in \mathbb{R}$, $|\nabla f|(x)$ is the minimal slope at x defined as follows

$$\big|\nabla f(x)\big| = \min\big\{\big|(1-\theta) f'_-(x) + \theta f'_+(x)\big|; \theta \in [0,1]\big\},$$

denoting by f'_- and f'_+ the left and right derivatives of f (which are always well-defined for a convex function).

(iii) *There exist $D, l_o > 0$ such that μ satisfies $\overline{\mathrm{T}}^-(\alpha)$ and $\overline{\mathrm{T}}^+(\alpha)$. Namely, it holds*

$$\overline{\mathcal{T}}_\alpha(\mu|\nu) \leq H(\nu|\mu) \quad \text{and} \quad \overline{\mathcal{T}}_\alpha(\nu|\mu) \leq H(\nu|\mu) \quad \forall \nu \in \mathcal{P}_1(\mathbb{R}),$$

for the cost function α defined by

$$\alpha(u) = \begin{cases} \frac{u^2}{2D} & \text{if } |u| \leq l_o D, \\ l_o|u| - l_o^2 D/2 & \text{if } |u| > l_o D. \end{cases}$$

In particular, with $\kappa := 5480$ and $c := 1/(10\sqrt{2})$,

- (i) $\Rightarrow$ (iii) *with $D = \kappa h^2$ and $l_o = c/h$,*
- (iii) $\Rightarrow$ (ii) *with $C = 2D$,*
- (i) $\Rightarrow$ (ii) *with $C = \kappa h^2$.*

The equivalence (i) $\Leftrightarrow$ (ii) goes back to Bobkov and Götze [8]. Hence, Theorem 1.6 completes the picture by showing that (i) or (ii) also characterize the measures satisfying a weak transport-entropy inequality with a cost function which is quadratic near zero and then linear (like θ_1). The dependence between the constants in the implication

(ii) $\Rightarrow$ (i) is not given for technical reasons. Indeed, the proof relies on an argument from [8] that uses a non trivial proof from [10] where one loses the explicit dependence on the constants.

We indicate that during the preparation of this work, we learned that the characterization of the convex Poincaré inequality in terms of the convex (τ)-Property (which is equivalent to the transport-entropy inequalities of item (iii), by duality, as we shall recall in Lemma 4.1) was obtained by Feldheim, Marsiglietti, Nayar and Wang in a recent paper [18].

The proof of Theorem 1.6 is given in Section 5. It uses results of independent interest like a new discrete logarithmic Sobolev inequality for the exponential measure τ (Lemma 5.2). By transportation techniques, such a logarithmic-Sobolev inequality provides logarithmic-Sobolev inequalities restricted to the class of convex or concave functions for measures satisfying the condition in item (i) (Proposition 5.1). Then the weak transport-entropy inequalities of item (iii) are obtained in their dual forms, involving infimum convolution operators (Lemma 4.1), by means of the Hamilton–Jacobi semi-group approach of Bobkov, Gentil and Ledoux [7], an approach also generalized in [24,25,32].

After the preparation of this paper, several additional progresses have been made in the understanding of weak transport inequalities and related functional inequalities [1,41,42,44]. We think it is worth mentioning these recent papers in some detail. In [1], Adamczak and Strzelecki have generalized the equivalence between (ii) and (iii) in our Theorem 1.6 to any dimension. It seems however, that for technical reasons this extension involves constants depending on the dimension. In [42], the fourth named author and Strzelecki have shown that, under some mild assumptions on the cost function θ, the weak transport inequality $\overline{\mathrm{T}}(\theta)$ (in dimension 1) is in fact equivalent to a modified log-Sobolev inequality restricted to convex functions. In [44], Strzelecka, Strzelecki and Tkocz have obtained weak transport inequalities with an optimal cost function θ for probability measures on the real line having log-concave tails. Finally, in [41], the fourth named author gave a simpler proof of the additivity formula (7) based on the dual formulation of weak transport costs obtained in [25].

The paper is organized as follows. In the next section, we introduce and recall some known properties of the convex ordering that we use in Section 3 to prove Theorem 1.5. Then, in Section 4, we very briefly recall the dual formulation of the weak-transport entropy inequalities $\overline{\mathrm{T}}^{\pm}$ and $\overline{\mathrm{T}}$, borowed from [25], which will be useful later on. Finally, Section 5 and 6 are devoted to the proofs of Theorem 1.6 and 1.3, respectively.

2. Convex ordering and a majorization theorem

This section is devoted to the study of the convex ordering. After recalling some classical definitions and results, we prove a majorization theorem which will be a key ingredient in the proof of Theorem 1.5.

2.1. *Convex ordering of probability measures*

We collect here some basic facts about convex ordering of probability measures. We refer the interested reader to [33] and [28] for further results and bibliographic references. The proofs are well-known, but we state some of them for completeness.

We start with the definition of the convex order.

Definition 2.1 (Convex order). Given $\nu_1, \nu_2 \in \mathcal{P}_1(\mathbb{R})$, we say that ν_2 dominates ν_1 in the convex order, and write $\nu_1 \preceq \nu_2$, if for all convex functions f on $\mathbb{R}$, $\int_{\mathbb{R}} f\, d\nu_1 \leq \int_{\mathbb{R}} f\, d\nu_2$.

Remark 2.2. Observe that for any probability measure belonging to $\mathcal{P}_1(\mathbb{R})$ the integral of any convex function always makes sense in $\mathbb{R} \cup \{+\infty\}$.

The convex ordering of probability measures can be determined by testing only some restricted classes of convex functions as the following proposition indicates.

Proposition 2.3. *Given $\nu_1, \nu_2 \in \mathcal{P}_1(\mathbb{R})$, the following are equivalent*:

(i) $\nu_1 \preceq \nu_2$.

(ii) $\int x\nu_1(dx) = \int x\nu_2(dx)$ *and for all Lipschitz, non-decreasing and non-negative convex functions* $f : \mathbb{R} \to \mathbb{R}^+$, $\int f(x)\nu_1(dx) \le \int f(x)\nu_2(dx)$.
(iii) $\int x\nu_1(dx) = \int x\nu_2(dx)$ *and for all* $t \in \mathbb{R}$, $\int [x-t]_+\nu_1(dx) \le \int [x-t]_+\nu_2(dx)$.

For the reader's convenience and for the sake of completeness, we sketch the proof of this classical result. We refer to [33] for more details.

Sketch of the proof. Let us show that (i) is equivalent to (ii). First, since the functions $x \mapsto x$ and $x \mapsto -x$ are both convex, it is clear that $\nu_1 \preceq \nu_2$ implies $\int x\nu_1(dx) = \int x\nu_2(dx)$ so that (i) implies (ii). Conversely, since the graph of a convex function always lies above its tangent, subtracting an affine function if necessary, one can restrict to non-negative convex functions. Moreover, if $f : \mathbb{R} \to \mathbb{R}^+$ is a convex function, then $f_n : \mathbb{R} \to \mathbb{R}^+$ defined by $f_n = f$ on $[-n, n]$, $f_n(x) = f_n(n) + f_n'(n)(x-n)$ if $x \ge n$ and $f_n(x) = f_n(-n) + f_n'(-n)(x+n)$ if $x \le -n$ (where f_n' denotes the right derivative of f) is Lipschitz and converges monotonically to f as n goes to infinity. The monotone convergence theorem then shows that one can further restrict to *Lipschitz* convex functions. Finally, up to the subtraction of an affine map, any Lipschitz convex function is non-decreasing, proving that (ii) implies (i).

Now it is not difficult to check that any convex, non-decreasing Lipschitz function $f : \mathbb{R} \to \mathbb{R}^+$ can be approached by a non-increasing sequence of functions of the form $\alpha_0 + \sum_{i=1}^n \alpha_i [x - t_i]_+$, with $\alpha_i \ge 0$ and $t_i \in \mathbb{R}$. This shows that (ii) and (iii) are equivalent. □

The next classical result, due to Strassen [43], characterizes the convex ordering in terms of martingales. Recall that a couple (X, Y) of integrable random variables is a martingale if $\mathbb{E}[Y|X] = X$ almost surely.

Theorem 2.4 (Strassen). *Given* $\nu_1, \nu_2 \in \mathcal{P}_1(\mathbb{R})$, *the following are equivalent*:

(i) $\nu_1 \preceq \nu_2$.
(ii) *There exists a martingale* (X, Y) *such that* X *has law* ν_1 *and* Y *has law* ν_2.

We refer to [25] for a (two-line) proof of Theorem 2.4 involving Kantorovich duality for transport costs of the form $\overline{T}$.

2.2. *Majorization of vectors*

The convex ordering is closely related to the notion of majorization of vectors that we recall in the following definition. As for the previous subsection, all the proofs are well-known and we state them for completeness.

Definition 2.5 (Majorization of vectors). Let $a, b \in \mathbb{R}^n$; one says that a is *majorized* by b, if the sum of the largest j components of a is less than or equal to the corresponding sum of b, for every j, and if the total sum of the components of both vectors are equal.

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

$$a_n + a_{n-1} + \cdots + a_{n-j+1} \le b_n + b_{n-1} + \cdots + b_{n-j+1}, \quad \text{for } j = 1, \dots, n-1,$$

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

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

Proposition 2.6. *Let* $a, b \in \mathbb{R}^n$ *and set* $\nu_1 = \frac{1}{n}\sum_{i=1}^n \delta_{a_i}$ *and* $\nu_2 = \frac{1}{n}\sum_{i=1}^n \delta_{b_i}$. *The following are equivalent*:

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

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

Proof. Assume without loss of generality that the components of a and b are sorted in increasing order. We observe first that, by construction, the equality $\int x\nu_1(dx) = \int x\nu_2(x)$ is equivalent to $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i$.

We will first prove that (i) implies (ii). By item (iii) of Proposition 2.3 we only need to prove that $a \preceq b$ implies

$$\sum_{k=1}^n [a_k - t]_+ \le \sum_{k=1}^n [b_k - t]_+, \quad \forall t \in \mathbb{R}. \tag{8}$$

Assume that $t \le \max a_k$ (otherwise (8) obviously holds). Then, let k_o be the smallest k such that $a_k \ge t$ so that $\sum_{k=1}^n [a_k - t]_+ = \sum_{k=k_o}^n (a_k - t)$. Therefore, by the majorization assumption (which guarantees that $\sum_{k=k_o}^n a_k \le \sum_{k=k_o}^n b_k$), we get

$$\sum_{k=1}^n [a_k - t]_+ = \sum_{k=k_o}^n (a_k - t) \le \sum_{k=k_o}^n b_k - t \le \sum_{k=1}^n [b_k - t]_+.$$

Conversely, let us prove that (ii) implies (i). Fix $k \in \{1, \dots, n\}$ and set $f_k(x) := [x - b_k]_+$, $x \in \mathbb{R}$. Plugging f_k into item (ii) of Proposition (2.3) leads to

$$\begin{aligned}\sum_{i=k}^n a_i - b_k \le \sum_{i=1}^n [a_i - b_k]_+ &= n \int f_k(x)\nu_1(dx) \\ &\le n \int f_k(x)\nu_2(dx) = \sum_{i=1}^n [b_i - b_k]_+ = \sum_{i=k}^n b_i - b_k,\end{aligned}$$

so that $\sum_{i=k}^n a_i \le \sum_{i=k}^n b_i$, which proves that a is majorized by b. □

Next we recall a simple classical consequence of Proposition 2.6 in terms of discrete optimal transport on the line.

Proposition 2.7. *Let $x, y \in \mathbb{R}^n$ be two vectors whose coordinates are listed in non-decreasing order (i.e., $x_1 \le x_2 \le \cdots \le x_n$, $y_1 \le y_2 \le \cdots \le y_n$). Then for all permutation σ of $\{1, \dots, n\}$ and all convex functions $\theta : \mathbb{R} \to \mathbb{R}$, it holds*

$$\sum_{i=1}^n \theta(x_i - y_i) \le \sum_{i=1}^n \theta(x_i - y_{\sigma(i)}).$$

Proof. Since, for all k, $\sum_{i=k}^n y_i \ge \sum_{i=k}^n y_{\sigma(i)}$, it holds for $\sum_{i=k}^n (x_i - y_i) \le \sum_{i=k}^n (x_i - y_{\sigma(i)})$ (with equality for $k = 1$). Therefore, denoting $y_\sigma = (y_{\sigma(1)}, \dots, y_{\sigma(n)})$, it holds $x - y \preceq x - y_\sigma$. Applying Proposition 2.6 completes the proof. □

Remark 2.8. In particular, let μ, ν are two discrete probability measures on $\mathbb{R}$ of the form

$$\mu = \frac{1}{n}\sum_{i=1}^n \delta_{x_i} \quad \text{and} \quad \nu = \frac{1}{n}\sum_{i=1}^n \delta_{y_i},$$

where the x_i's and the y_i's are in increasing order, and assume for simplicity that the x_i's are distinct. Then the map T sending x_i on y_i for all i realizes the optimal transport of μ onto ν for every cost function θ.

We end this section with a characterization of the convex ordering (or equivalently of the majorization of vectors, thanks to Proposition 2.6), due to Hardy–Littlewood–Pólya [26] ((i) ⇔ (ii)) and Rado [39] ((i) ⇔ (iii)). For simplicity, we denote by $\mathcal{S}_n$ the set of all permutations of $\{1, 2, \dots, n\}$ and, given $\sigma \in \mathcal{S}_n$ and $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, we set $x_\sigma := (x_{\sigma(1)}, \dots, x_{\sigma(n)})$.

Theorem 2.9 (Hardy–Littlewood–Pólya, Rado). *Given $a, b \in \mathbb{R}^n$, the following are equivalent*:

(i) *The vector a is majorized by b.*
(ii) *There exists a doubly stochastic matrix P such that $a = bP$.*
(iii) *There exists a collection of non-negative numbers $(\lambda_\sigma)_{\sigma\in\mathcal{S}_n}$ with $\sum_{\sigma\in\mathcal{S}_n}\lambda_\sigma = 1$ such that $a = \sum_{\sigma\in\mathcal{S}_n}\lambda_\sigma b_\sigma$ (i.e., a lies in the convex hull of the permutations of b).*

Below, we sketch a proof based on Strassen's Theorem and refer to [33, Theorem B.2, Chapter 2] for more classical arguments.

Proof. First we will prove that (i) implies (ii). According to Proposition 2.6, $a \preceq b$ is equivalent to saying that $\nu_1 = \frac{1}{n}\sum_{i=1}^n \delta_{a_i}$ is dominated by $\nu_2 = \frac{1}{n}\sum_{i=1}^n \delta_{b_i}$ in the convex order. Set $\mathcal{X} := \{a_1, \dots, a_n\}$, $\mathcal{Y} := \{b_1, \dots, b_n\}$, $k_x := \#\{i \in \{1, \dots, n\} : a_i = x\}$, $x \in \mathcal{X}$ and $\ell_y = \#\{i \in \{1, \dots, n\} : b_i = y\}$, $y \in \mathcal{Y}$ (where # denotes the cardinality); observe that $\nu_1 = \frac{1}{n}\sum_{x\in\mathcal{X}} k_x\delta_x$ and $\nu_2 = \frac{1}{n}\sum_{y\in\mathcal{Y}}\ell_y\delta_y$. According to the Strassen Theorem (Theorem 2.4), there exists a couple of random variables (X, Y) on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ such that X is distributed according to ν_1, and Y according to ν_2 and $X = \mathbb{E}[Y|X]$. Since X is a discrete random variable,

$$\mathbb{E}[Y|X] = \sum_{x\in\mathcal{X}} \frac{\mathbb{E}[Y\mathbf{1}_{X=x}]}{\mathbb{P}(X=x)}\mathbf{1}_{X=x}, \quad \text{a.s.}$$

Therefore, for all $x \in \mathcal{X}$,

$$x = \frac{\mathbb{E}[Y\mathbf{1}_{X=x}]}{\mathbb{P}(X=x)} = \sum_{y\in\mathcal{Y}} \ell_y y K_{y,x},$$

where $K_{y,x} := n\frac{\mathbb{P}(X=x,Y=y)}{k_x\ell_y}$. Hence $a = bP$ with $P_{j,i} := K_{b_j,a_i}$, $i, j = 1, \dots, n$. This proves item (ii), since P is doubly stochastic by construction.

If $a = bP$ with a doubly stochastic matrix P, then it is easily checked that $\sum_{i=1}^n f(a_i) \le \sum_{i=1}^n f(b_i)$ for any convex function f on $\mathbb{R}$ so that (ii) implies (i).

Finally, according to Birkhoff's theorem [33, Theorem A.2, Chapter 2], the extremes points of the set of doubly stochastic matrices are *permutation matrices*. Therefore every doubly stochastic matrix can be written as a convex combination of permutation matrices showing that (ii) and (iii) are equivalent. □

2.3. *Geometric aspects of convex ordering and a majorization theorem*

Contrary to the previous subsections, the results presented here are new. Fix some vector $b = (b_1, b_2, \dots, b_n)$ of $\mathbb{R}^n$ with *distinct* components (for simplicity). We will be working with the convex hull of the permutations of b, a polytope we denote by $\mathrm{Perm}(b)$ and defined as

$$\mathrm{Perm}(b) := \left\{ \sum_{\sigma\in\mathcal{S}_n} \lambda_\sigma b_\sigma, \text{ with } \lambda_\sigma \ge 0 \text{ and } \sum_{\sigma\in S_n}\lambda_\sigma = 1 \right\}.$$

Such a polytope is often refered to as the *Permutahedron* generated by b. According to Theorem 2.9, $\mathrm{Perm}(b) = \{a \in \mathbb{R}^n : a \preceq b\}$. Hence, $\mathrm{Perm}(b)$ is a subset of the following affine hyperplane

$$\mathcal{E}_b := \left\{ x \in \mathbb{R}^n : \sum_{i=1}^n x_i = \sum_{i=1}^n b_i \right\} = b + \mathcal{E}_0,$$

with $\mathcal{E}_0 := \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 0\}$.

We will be interested in the faces, facets containing a given face, and normal vectors to such facets of $\mathrm{Perm}(b)$. To this aim, we need to introduce some notations.

Denote by $[n]$ the set of integers from 1 to n. For $S \subset [n]$, let $v_S(b)$ denote the vector with the $|S|$ largest components of b in the positions indexed by S (in decreasing order, say), and the remaining $n - |S|$ lowest components of b in the other positions indexed by $[n] \setminus S$ (also in a decreasing order). Also, when $S \neq \varnothing$, we denote by $P_S(b)$ the set that contains the vector $v_S(b)$ along with all vectors obtained by permuting any subset of coordinates of $v_S(b)$, as long as the subset is contained in S or in $[n] \setminus S$. (That is, the only permutations that are *not* allowed are those that involve elements from *both* S and $[n] \setminus S$). More precisely

$$P_S(b) := \big\{ \big(v_S(b)\big)_\sigma, \sigma \in \mathcal{S}_n \text{ such that } \sigma(S) = S \big\},$$

where $\sigma(\mathcal{S}) := \{\sigma(i), i \in \mathcal{S}\}$ denotes the image of $\mathcal{S}$ by σ.

More generally, given a partition $\mathcal{S} = (S_1, S_2, \ldots, S_k)$ of $[n]$, let $v_{\mathcal{S}}(b)$ denote the vector with the largest $|S_1|$ coordinates of b in the positions indexed by S_1 (in decreasing order), then the next largest $|S_2|$ coordinates in the positions indexed by S_2 and so on (as an illustration, for $b = (1, 4, 5, -2, 3, 9, 6, -5) \in \mathbb{R}^8$ and $\mathcal{S} = (S_1, S_2, S_3)$ with $S_1 = \{1, 2\}$, $S_2 = \{3, 6, 7\}$ and $S_3 = \{4, 5, 8\}$, we get $v_{\mathcal{S}}(b) = (9, 6, \mathbf{5}, 1, -2, \mathbf{4}, \mathbf{3}, -5)$ where the italic positions refer to the set S_1, the bold positions to the set S_2 and the remaining positions to S_3). Also, we denote by $P_{\mathcal{S}}(b)$ the set containing the vector $v_{\mathcal{S}}(b)$ along with all vectors obtained by permuting the coordinates of $v_{\mathcal{S}}(b)$ that belong to the same S_i:

$$P_{\mathcal{S}}(b) := \big\{ \big(v_S(b)\big)_\sigma, \sigma \in \mathcal{S}_n \text{ such that for all } i, \sigma(S_i) = S_i \big\}.$$

Now we recall two geometric definitions/facts from [6].

Fact 1: A *facet* of $\mathrm{Perm}(b)$ is the convex hull of $P_S(b)$, for some $S \neq \varnothing, [n]$.

Fact 2: A *face* of $\mathrm{Perm}(b)$ is the convex hull of $P_{\mathcal{S}}(b)$, for some partition $\mathcal{S} = (S_1, S_2, \ldots, S_k)$ of $[n]$ with $k \geq 3$. Furthermore, given a face $F = \mathrm{Conv}(P_{\mathcal{S}}(b))$, there exist exactly $k - 1$ facets containing F that are obtained by coalescing the first and last several S_i's in $\mathcal{S}$: that is, for each $1 \leq j \leq k - 1$, the facet F_j containing F can be described by taking the partition $[n] = T_1 \cup T_2$ with $T_1 = S_1 \cup \cdots \cup S_j$, and $T_2 = S_{j+1} \cup \cdots \cup S_k$.

The next theorem, which we may call the Majorization Theorem, is a key ingredient in the proof of Theorem 1.5. It provides a geometric interpretation of majorization in terms of projection.

Theorem 2.10 (Majorization Theorem). *Let $a, b \in \mathbb{R}^n$, assume that b has distinct coordinates and that $a \notin \mathrm{Perm}(b)$. Then the following are equivalent*:

(i) $\hat{c} \in \mathrm{Perm}(b)$ *satisfies*

$$a - \hat{c} \preceq a - c, \quad \forall c \in \mathrm{Perm}(b);$$

(ii) $\hat{c}$ *is the closest point of* $\mathrm{Perm}(b)$ *to* a; *that is*,

$$\hat{c} := \arg \min_{c \in \mathrm{Perm}(b)} \big(\|a - c\|_2 \big).$$

Moreover the vector $\hat{c}$ is sorted as $a : (a_i \leq a_j) \Rightarrow (\hat{c}_i \leq \hat{c}_j)$, for all i, j.

Let us recall that the orthogonal projection of a point a on the polytope $\mathrm{Perm}(b)$ is the unique $\bar{c} \in \mathrm{Perm}(b)$ such that

$$\langle a - \bar{c}, c - \bar{c} \rangle \leq 0, \quad \forall c \in \mathrm{Perm}(b). \tag{9}$$

Proof. Observe that if $\sum_{i=1}^n a_i \neq \sum_{i=1}^n b_i$, then letting $\tilde{a} := a - \frac{k}{n}(1, 1, \ldots, 1)$ with $k := \sum_{i=1}^n a_i - \sum_{i=1}^n b_i$, we see (using (9)) that the orthogonal projection of a and $\tilde{a}$ on $\mathrm{Perm}(b)$ are equal (to some point we denote by $\hat{c}$, say), and that $a - \hat{c} \preceq a - c$ if and only if $\tilde{a} - \hat{c} \preceq \tilde{a} - c$. Therefore we can assume without loss that a and b are such that $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i$.

First let us prove that (i) implies (ii) which is the easy part of the proof. Let $\bar{c}$ be the closest point of $\mathrm{Perm}(b)$ to a (i.e., $\bar{c} := \arg\min_{c\in\mathrm{Perm}(b)}(\|a-c\|_2)$). Then by (i), $a-\hat{c} \preceq a-\bar{c}$, which, by Proposition 2.6 (applied to $f(x)=x^2$) implies that $\sum_{i=1}^n (a-\hat{c})_i^2 \le \sum_{i=1}^n (a-\bar{c})_i^2$. By definition of $\bar{c}$, this is possible only if $\hat{c}=\bar{c}$.

Next let us prove that (ii) implies (i). For the sake of clarity, we first deal with the simple case when $\hat{c}$ lies on a *facet* of $\mathrm{Perm}(b)$, before dealing with the general case of $\hat{c}$ being on a face.

Let $\hat{c}$ be the closest point of $\mathrm{Perm}(b)$ to a. Since $\mathrm{Perm}(b)$ is invariant by permutation, it easily follows from Proposition 2.7 that the coordinates of $\hat{c}$ are in the same order as the coordinates of a. Hence, we are left with the proof that $a-\hat{c} \preceq a-c$ for all $c\in\mathrm{Perm}(b)$.

(a) A simple case: $\hat{c}\in F$ for some facet F. Since $\hat{c}$ is chosen from $\mathrm{Perm}(b)$, and since we assumed that $\sum_i b_i = \sum_i a_i$, we have $\sum_i (a-\hat{c})_i = 0$. Writing $\alpha := a-\hat{c} \in \mathcal{E}_0$, suppose that α is perpendicular to the affine subspace $\mathcal{H} := \mathcal{H}_F$ containing a facet F, defined by some nonempty subset S of $[n]$. For all $x, y \in F$, we thus have $\langle \alpha, x-y\rangle = 0$. Choosing $x = v_S(b)$ and $y = x_{\tau_{ij}}$ obtained by permuting two coordinates of x whose indices are both in S or both in S^c (i.e., $\tau_{ij}=(ij)$ is the transposition that permutes i and j, with $i,j\in S$, or $i,j\in S^c$), one sees that the coordinates of α are constant on S and S^c. We denote by α_S and α_{S^c} the values of α on these sets, which verify $k\alpha_S + (n-k)\alpha_{S^c} = 0$ since $\alpha\in\mathcal{E}_0$.

Now (recalling that $\alpha = a-\hat{c}$) our task is to show that

$$\alpha \preceq \alpha - \big(c'-\hat{c}\big), \quad \text{for every } c'\in\mathrm{Perm}(b).$$

This amounts to showing that

$$\alpha \preceq \alpha - c, \quad \text{for every } c \text{ such that } \langle \alpha, c\rangle \le 0, \quad \text{and} \quad \sum_i c_i = 0.$$

Indeed, on the one hand, by (9), the choice of $\hat{c}$ implies that for every $c'\in\mathrm{Perm}(b)$, we have $\langle \alpha, c'-\hat{c}\rangle \le 0$, and on the other hand, since $\hat{c}, c' \in \mathrm{Perm}(b)$, necessarily $\sum_i c_i = \sum_i c'_i - \sum_i \hat{c}_i = 0$.

Now $\langle \alpha, c\rangle \le 0$ and $\sum_i c_i = 0$ together imply (recalling that α is constant on S and S^c),

$$(\alpha_S - \alpha_{S^c}) \sum_{i\in S} c_i \le 0.$$

Let us assume that $\alpha_S > \alpha_{S^c}$. Then denoting by $c_S = \sum_{i\in S} c_i$ and by $c_{S^c} = \sum_{i\in S^c} c_i$, one has $c_S \le 0$ and $c_{S^c} \ge 0$. Therefore, for any convex function f on $\mathbb{R}$, according to Jensen's inequality and by convexity, we get

$$\begin{aligned}
\sum_{i=1}^n f(\alpha_i - c_i) &= k\frac{\sum_{i\in S} f(\alpha_S - c_i)}{k} + (n-k)\frac{\sum_{i\in S^c} f(\alpha_{S^c} - c_i)}{n-k} \\
&\ge k f\left(\alpha_S - \frac{c_S}{k}\right) + (n-k) f\left(\alpha_{S^c} - \frac{c_{S^c}}{n-k}\right) \\
&\ge k f(\alpha_S) + (n-k) f(\alpha_{S^c}) - f'(\alpha_S) c_S - f'(\alpha_{S^c}) c_{S^c} \\
&\ge \sum_{i=1}^n f(\alpha_i),
\end{aligned}$$

where the last inequality comes from the fact that $f'(\alpha_S)c_S + f'(\alpha_S^c)c_{S^c} = c_S(f'(\alpha_S) - f'(\alpha_{S^c})) \le 0$. According to Proposition 2.6, we conclude that $\alpha \preceq \alpha - c$ which is the expected result.

(b) The general case. Suppose that $\hat{c}$ lies in a face F of the polytope. This face is related to a partition $\mathcal{S} = (S_1, \dots, S_k)$ of $[n]$, with $k\ge 3$. Then $\alpha := a-\hat{c} \in N(F)$, where $N(F)$ denotes the normal cone of F. Recall that the extreme rays of $N(F)$ are given by the facet directions for the facets containing F. For all $i\in\{1,\dots,n-1\}$, let us denote by F_i the facet containing F associated to the partition $\mathcal{T}_i = \{S_1\cup\cdots\cup S_i; S_{i+1}\cup\cdots\cup S_k\}$, $1\le i\le k-1$. Consider the vectors $p_1, p_2, \dots, p_{k-1} \in \mathcal{E}_0$ defined by

$$p_i = \mathbf{1}_{S_1\cup S_2\cup\cdots\cup S_i} - \frac{k_i}{n}\mathbf{1}_{[n]},$$

where $\mathbf{1}_T$ denotes the $0-1$ indicator vector of T, for $T \subseteq [n]$, and $k_i = |S_1| + \cdots + |S_i|$. For each i, the vector p_i is orthogonal to the facet F_i. Moreover, for all $c \in \mathrm{Perm}(b)$ one may check that $\langle c, p_i \rangle \leq \langle v_{\mathcal{T}_i}, p_i \rangle$, with equality on F_i. This shows that p_i is an outward normal vector to F_i. Therefore $N(F)$ is the conical hull of the p_i's, and so we may express α, for a suitable choice of $\lambda_i \geq 0$, as:

$$\alpha = \sum_i \lambda_i \mathbf{1}_{S_1 \cup S_2 \cup \cdots \cup S_i} - \sigma \mathbf{1}_{[n]},$$

where $\sigma = (1/n)[\sum_{i=1}^{k-1} \lambda_i |S_1| + \sum_{i=2}^{k-1} \lambda_i |S_2| + \cdots + \lambda_{k-1}|S_{k-1}|]$. In particular, α is constant on each S_j: for all $i \in S_j$, $\alpha_i = (\sum_{p=j}^{k-1} \lambda_p) - \sigma := A_j$.

In order to establish (i), we need to show that

$$\alpha \preceq \alpha - (c - \hat{c}), \quad \forall c \in \mathrm{Perm}(b),$$

or in other words, we need to show that

$$\alpha \preceq \alpha - c', \quad \forall c' \in \mathrm{Perm}(b) - \hat{c}.$$

We now use again the fact that our choice of $\hat{c}$ implies that, for all $1 \leq i \leq k-1$,

$$\langle p_i, \hat{c} \rangle \geq \langle p_i, c \rangle, \quad \forall c \in \mathrm{Perm}(b).$$

This in turn gives the following:

$$\mathrm{Perm}(b) - \hat{c} \subseteq \left\{ c' : \langle c', p_i \rangle \leq 0, \forall i \right\}.$$

Thus using $N(F)^0 := \{d \in \mathcal{E}_0;\ \langle d, p_i \rangle \leq 0, \forall i\}$ to denote the *polar cone*, it then suffices to show that for α (as above),

$$\alpha \preceq \alpha - d, \quad \forall d \in N(F)^0.$$

Now, $d \in N(F)^0$ implies that

$$\langle d, \mathbf{1}_{S_1 \cup S_2 \cup \cdots \cup S_j} \rangle \leq 0 \quad \text{and} \quad \sum_i d_i = 0,$$

therefore denoting $E_j = \sum_{i \in S_1 \cup \cdots \cup S_j} d_i$, for all $j \in \{0, 1, \ldots, k\}$, one has $E_j \leq 0$ and $E_0 = E_k = 0$.

Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function; denoting by f' its right derivative, the convexity of f implies that

$$\sum_{i=1}^n f(\alpha_i - d_i) = \sum_{j=1}^k \sum_{i \in S_j} f(A_j - d_i) \geq \sum_{j=1}^k |S_j| f(A_j) - \sum_{j=1}^k f'(A_j) D_j,$$

where $D_j = \sum_{i \in S_j} d_i$. Now, using an Abel transform (and the fact that $E_0 = E_k = 0$), one gets

$$\sum_{j=1}^k f'(A_j) D_j = \sum_{j=1}^k f'(A_j)(E_j - E_{j-1}) = \sum_{j=1}^{k-1} \big(f'(A_j) - f'(A_{j+1}) \big) E_j \leq 0,$$

where the inequality comes from $E_j \leq 0$, $A_j \geq A_{j+1}$ and the monotonicity of f'. Therefore, one gets

$$\sum_{i=1}^n f(\alpha_i - d_i) \geq \sum_{j=1}^k |S_j| f(A_j) = \sum_{i=1}^n f(a_i),$$

which proves that $a \preceq a - d$, thanks to Proposition 2.6, as expected. This completes the proof. □

3. Properties of the optimal coupling for weak transport costs

This section is devoted to the proof of Theorem 1.5. We will establish first a preliminary result which gives some connection between $\mathcal{T}$ and $\overline{\mathcal{T}}$. In the sequel, we denote by $\mathrm{Im}(\mu)$, respectively $\mathrm{Im}^{\uparrow}(\mu)$, the set of probability measures on $\mathbb{R}$ which are images of μ under some map $S:\mathbb{R}\to\mathbb{R}$, respectively some non-decreasing map S, i.e.,

$$\mathrm{Im}(\mu)=\big\{\gamma\in\mathcal{P}(\mathbb{R}):\exists S:\mathbb{R}\to\mathbb{R}\text{ measurable such that }\gamma=S_{\#}\mu\big\},$$

and

$$\mathrm{Im}^{\uparrow}(\mu)=\big\{\gamma\in\mathcal{P}(\mathbb{R}):\exists S:\mathbb{R}\to\mathbb{R}\text{ non-decreasing such that }\gamma=S_{\#}\mu\big\}. \tag{10}$$

Proposition 3.1. *For all probability measures $\mu,\nu\in\mathcal{P}_1(\mathbb{R})$ and all convex function $\theta:\mathbb{R}^+\to\mathbb{R}^+$, it holds*

$$\inf_{\gamma\preceq\nu,\gamma\in\mathrm{Im}^{\uparrow}(\mu)}\mathcal{T}_\theta(\gamma,\mu)\geq\overline{\mathcal{T}}_\theta(\nu|\mu)\geq\inf_{\gamma\preceq\nu,\gamma\in\mathrm{Im}(\mu)}\mathcal{T}_\theta(\gamma,\mu).$$

Remark 3.2. Note that when μ has no atoms, then $\mathrm{Im}^{\uparrow}(\mu)=\mathrm{Im}(\mu)$. If μ is a discrete probability measure, then the two sets may be different. For instance, if $\mu=\frac{1}{3}\delta_0+\frac{2}{3}\delta_1$, then $\gamma=\frac{2}{3}\delta_0+\frac{1}{3}\delta_1$ is in $\mathrm{Im}(\mu)$ but not in $\mathrm{Im}^{\uparrow}(\mu)$. In the proof of Theorem 1.5 below, we will use Proposition 3.1 with μ being the uniform distribution on n distinct points for which it is clear that $\mathrm{Im}^{\uparrow}(\mu)=\mathrm{Im}(\mu)$.

Proof. First we will prove that $\overline{\mathcal{T}}_\theta(\nu|\mu)\geq\inf_{\gamma\preceq\nu,\gamma\in\mathrm{Im}(\mu)}\mathcal{T}_\theta(\gamma,\mu)$. To that aim, denote by $\pi(dx\,dy)=p(x,dy)\mu(dx)$ some coupling between μ and ν and set $S(x):=\int yp(x,dy)$, $x\in\mathbb{R}$. Clearly $S_{\#}\mu\in\mathrm{Im}(\mu)$. Moreover if $f:\mathbb{R}\to\mathbb{R}$ is some convex function, by Jensen's inequality, it holds

$$\int f(x)S_{\#}\mu(dx)=\int f\bigg(\int yp(x,dy)\bigg)\mu(dx)\leq\iint f(y)p(x,dy)\mu(dx)=\int f(y)\nu(dy)$$

so that $S_{\#}\mu\preceq\nu$. Therefore,

$$\int\theta\bigg(x-\int yp(x,dy)\bigg)\mu(dx)=\int\theta\big(x-S(x)\big)\mu(dx)\geq\mathcal{T}_\theta(S_{\#}\mu,\mu)\geq\inf_{\gamma\preceq\nu,\gamma\in\mathrm{Im}(\mu)}\mathcal{T}_\theta(\gamma,\mu),$$

from which the claim follows by taking the infimum over p.

Now we turn to the proof of the inequality $\overline{\mathcal{T}}_\theta(\nu|\mu)\leq\inf_{\gamma\preceq\nu,\gamma\in\mathrm{Im}^{\uparrow}(\mu)}\mathcal{T}_\theta(\gamma,\mu)$. Assume that $\gamma\preceq\nu$ and that $\gamma=S_{\#}\mu$ for some non-decreasing map S. According to Strassen's theorem, there exists a coupling π_1 with first marginal γ and second marginal ν such that $\pi_1(dx\,dy)=p_1(x,dy)\gamma(dx)$ and $x=\int_{\mathbb{R}}yp_1(x,dy)$, γ almost everywhere. For all $x\in\mathbb{R}$, define the following probability measure $p(x,dy):=p_1(S(x),dy)$. Then for all bounded continuous function f, it holds

$$\begin{aligned}\iint f(y)p(x,dy)\mu(dx)&=\iint f(y)p_1\big(S(x),dy\big)\mu(dx)\\&=\iint f(y)p_1(x,dy)\gamma(dx)=\int f(y)\nu(dy).\end{aligned}$$

Thus the coupling $\pi(dx\,dy)=p(x,dy)\mu(dx)$ has μ as first marginal and ν as second marginal. Moreover, by definition of p_1 and p, μ almost everywhere, it holds

$$\int yp(x,dy)=\int yp_1\big(S(x),dy\big)=S(x).$$

Since S is non-decreasing, it realizes the optimal transport between μ and ν for the classical transport cost $\mathcal{T}_\theta$. This last property is a classical result of optimal transport. Let us give a quick justification. Let $\Gamma=\{(x,S(x));x\in\mathbb{R}\}$.

If $(x_i, y_i)_{1\le i\le n}$ is a finite family of points of Γ, then according to Proposition 2.7, it holds $\sum_{i=1}^n \theta(x_i - y_i) \le \sum_{i=1}^n \theta(x_i - y_{i+1})$ (denoting $y_{n+1} = y_1$). Therefore Γ is c-cyclically monotone for the cost function $c(x, y) = \theta(x - y)$, $x, y \in \mathbb{R}$. Therefore, according to [48, Theorem 5.10 (ii)] the transport map S is optimal for the cost $\mathcal{T}_\theta$. So it follows that

$$\mathcal{T}_\theta(\gamma, \mu) = \int \theta\big(|x - S(x)|\big)\mu(dx) = \int \theta\left(\left|x - \int yp(x, dy)\right|\right)\mu(dx) \ge \overline{\mathcal{T}}_\theta(\nu|\mu),$$

which achieves the proof by taking the infimum over γ. □

We are now in a position to prove Theorem 1.5.

Proof of Theorem 1.5. The proof of the first part of Theorem 1.5 is divided into two steps. In the first step we will deal with uniform discrete measures on n points, while in the second step we will use an approximation argument in order to reach any measure.

Step 1. We first deal with

$$\mu := \frac{1}{n}\sum_{i=1}^n \delta_{a_i} \quad \text{and} \quad \nu := \frac{1}{n}\sum_{i=1}^n \delta_{b_i},$$

with $a_1 < a_2 < \cdots < a_n$ and $b_1 < b_2 < \cdots < b_n$. Set $a := (a_1, \dots, a_n)$ and $b := (b_1, \dots, b_n)$. According to Theorem 2.10, there exists some $\hat{c} \in \mathrm{Perm}(b)$ such that $a - \hat{c} \preceq a - c$, for all $c \in \mathrm{Perm}(b)$. Moreover the coordinates of $\hat{c}$ satisfy $\hat{c}_i \le \hat{c}_{i+1}$. Set $\hat{\gamma} := \frac{1}{n}\sum_{i=1}^n \delta_{\hat{c}_i}$ and observe that ν dominates $\hat{\gamma}$ in the convex order and $\hat{\gamma} \in \mathrm{Im}^\uparrow(\mu)$, where $\mathrm{Im}^\uparrow(\mu)$ is defined by (10).

Now for any $\gamma := \frac{1}{n}\sum_{i=1}^n \delta_{c_i} \in \mathrm{Im}^\uparrow(\mu)$ with $c_i \le c_{i+1}$ and for any convex cost function θ, it holds (since the coordinates are non-decreasing)

$$\mathcal{T}_\theta(\gamma, \mu) = \frac{1}{n}\sum_{i=1}^n \theta\big(|a_i - c_i|\big).$$

In particular

$$\mathcal{T}_\theta(\hat{\gamma}, \mu) = \frac{1}{n}\sum_{i=1}^n \theta\big(|a_i - \hat{c}_i|\big) \le \inf_{c\in\mathrm{Perm}(b)} \frac{1}{n}\sum_{i=1}^n \theta\big(|a_i - c_i|\big). \tag{11}$$

A probability γ such that $\gamma \preceq \nu$, $\gamma \in \mathrm{Im}^\uparrow(\mu)$ is of the form $\gamma = \frac{1}{n}\sum_{i=1}^n \delta_{c_i}$ with $c_i \le c_{i+1}$ and $c = (c_1, \dots, c_n) \in \mathrm{Perm}(b)$, and for such a c, it holds $\frac{1}{n}\sum_{i=1}^n \theta(|a_i - c_i|) = \mathcal{T}_\theta(\gamma, \mu)$. Therefore, the latter implies

$$\mathcal{T}_\theta(\hat{\gamma}, \mu) \le \inf_{\gamma\preceq\nu, \gamma\in\mathrm{Im}^\uparrow(\mu)} \mathcal{T}_\theta(\gamma, \mu) = \overline{\mathcal{T}}_\theta(\nu|\mu),$$

where the last equality follows from Proposition 3.1 and the fact that for such a distribution μ, it holds $\mathrm{Im}(\mu) = \mathrm{Im}^\uparrow(\mu)$ (see Remark 3.2). Since obviously $\overline{\mathcal{T}}_\theta(\nu|\mu) \le \mathcal{T}_\theta(\hat{\gamma}, \mu)$, we conclude that $\mathcal{T}_\theta(\hat{\gamma}, \mu) = \overline{\mathcal{T}}_\theta(\nu|\mu)$ as expected.

Step 2. In the second step we deal with the general case using an approximation argument.

Let μ and ν be two elements of $\mathcal{P}_1(\mathbb{R})$. By assumption, $\int |x|\mu(dx) < \infty$ and $\int |x|\nu(dx) < \infty$, hence, according to the de la Vallée-Poussin Theorem (see, e.g., [12, Theorem 4.5.9]), there exists an increasing convex function $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ such that $\beta(t)/t \to \infty$ as $t \to \infty$ and such that $\int \beta(|x|)\mu(dx) < \infty$ and $\int \beta(|x|)\nu(dx) < \infty$.

Next we will construct discrete approximations of μ and ν. According to Varadarajan's theorem (see, e.g., [17, Theorem 11.4.11]), if X_i is an i.i.d. sequence of law μ, then, with probability 1, the empirical measure $L_n^X := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ converges weakly to μ. On the other hand, according to the strong law of large numbers, with probability 1, $\frac{1}{n}\sum_{i=1}^n |X_i| \to \int |x|\mu(dx)$ as $n \to \infty$. Let us take $(x_i)_{i\ge1}$, a positive realization of these events and set $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i^{(n)}}$, where $x_1^{(n)} \le x_2^{(n)} \le \cdots \le x_n^{(n)}$ denotes the increasing re-ordering of the vector $(x_1, x_2, \dots, x_n)$.

Then the sequence μ_n converges weakly to μ and $\int |x|\mu_n(dx) \to \int |x|\mu(dx)$. According to [48, Theorem 6.9], this is equivalent to the convergence of the W_1 distance : $W_1(\mu_n, \mu) \to 0$ as $n \to \infty$. Note that one can assume that the points $x_i^{(n)}$ are distinct. Indeed, if this is not the case, then letting $\tilde{x}_i^{(n)} = x_i^{(n)} + i/n^2$ one obtains distinct points and it is not difficult to check that $\tilde{\mu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{\tilde{x}_i^{(n)}}$ still weakly converges to μ (for instance the W_1 distance between μ_n and $\tilde{\mu}_n$ is easily bounded from above by $(n+1)/(2n^2)$). The same argument yields a sequence $\nu_n = \frac{1}{n}\sum_{i=1}^n \delta_{y_i^{(n)}}$ with $y_i^{(n)} < y_{i+1}^{(n)}$ converging to ν in the W_1 sense. It is not difficult to check (invoking the strong law of large numbers again) that one can further impose that $\int \beta(|x|)\nu_n(dx) \to \int \beta(|x|)\nu(dx)$, as $n \to \infty$.

For all $n \geq 1$, one applies the result proved in the first step: there exists a unique probability measure $\hat{\gamma}_n \preceq \nu_n$ such that

$$\overline{\mathcal{T}}_\theta(\nu_n|\mu_n) = \mathcal{T}_\theta(\hat{\gamma}_n, \mu_n),$$

for all convex cost functions θ. Let us show that one can extract from $\hat{\gamma}_n$ a subsequence converging to some $\hat{\gamma}$ in $\mathcal{P}_1(\mathbb{R})$ for the W_1 distance. By construction $\int \beta(|x|)\nu_n(dx) \to \int \beta(|x|)\nu(dx)$ and so $M = \sup_{n\geq 1}\int \beta(|x|)\nu_n(dx)$ is finite. Since $\hat{\gamma}_n \preceq \nu_n$ and since the function $x \mapsto \beta(|x|)$ is convex, it thus holds $\int \beta(|x|)\gamma_n(dx) \leq \int \beta(|x|)\nu_n(dx) \leq M$. In particular, setting $c(R) = \inf_{t\geq R}\beta(t)/t$, $R > 0$, Markov's inequality easily implies that

$$\int_{[-R,R]^c} |x|\hat{\gamma}_n(dx) \leq \frac{\int \beta(|x|)\nu_n(dx)}{c(R)} \leq \frac{M}{c(R)}.$$

Consider $\tilde{\gamma}_n$ defined by $\frac{d\tilde{\gamma}_n}{d\hat{\gamma}_n}(x) = \frac{1+|x|}{\int 1+|x|\hat{\gamma}_n(dx)}$. Then it holds,

$$\sup_{n\geq 1}\tilde{\gamma}_n\big([-R,R]^c\big) \leq \frac{2M}{c(R)}, \quad \forall R \geq 1,$$

and so the sequence $\tilde{\gamma}_n$ is tight. Therefore, according to the Prokhorov Theorem, extracting a subsequence if necessary, one can assume that $\tilde{\gamma}_n$ converges to some $\tilde{\gamma}$ for the weak topology. Extracting yet another subsequence if necessary, one can also assume that $\int(1+|x|)\gamma_n(dx)$ converges to some number $Z > 0$. The weak convergence of $\tilde{\gamma}_n$ to $\tilde{\gamma}$ means that $\int \varphi\, d\tilde{\gamma}_n \to \int \varphi\, d\gamma$ for all bounded continuous φ, which means that

$$\int \big(1+|x|\big)\varphi(x)\hat{\gamma}_n(dx) \to \int \big(1+|x|\big)\varphi(x)\hat{\gamma}(dx),$$

where $\hat{\gamma}(dx) = \frac{Z}{1+|x|}\tilde{\gamma}(dx) \in \mathcal{P}_1(\mathbb{R})$. Invoking again [48, Theorem 6.9], this implies $\hat{\gamma}_n \to \hat{\gamma}$ as $n \to \infty$ for the W_1 distance.

Now we will check that $\hat{\gamma}$ is such that $\overline{\mathcal{T}_\theta}(\nu|\mu) = \mathcal{T}_\theta(\hat{\gamma}, \mu)$ for all convex cost functions $\theta : \mathbb{R}^+ \to \mathbb{R}^+$. First assume that θ is Lipschitz, and denote by L_θ its Lipschitz constant. According to [25, Theorem 2.11], the following Kantorovich duality formula holds

$$\overline{\mathcal{T}}_\theta(\nu_n|\mu_n) = \sup_\varphi\left\{\int Q_\theta\varphi(x)\nu_n(dx) - \int \varphi(y)\mu_n(dy)\right\},$$

where the supremum is taken over the set of convex functions φ bounded from below, with $Q_\theta\varphi(x) := \inf_{y\in\mathbb{R}}\{\varphi(y) + \theta(|x-y|)\}$, $x \in \mathbb{R}$. Define $\bar{\varphi}(y) := \sup_{x\in\mathbb{R}}\{Q_\theta\varphi(x) - \theta(|x-y|)\}$. Then it is easily checked that $\bar{\varphi} \leq \varphi$, $\bar{\varphi}$ is bounded from below and $Q_\theta\bar{\varphi} = Q_\theta\varphi$. Moreover, being a supremum of convex and L_θ-Lipschitz functions, the function $\bar{\varphi}$ is also convex and L_θ-Lipschitz. Therefore, the supremum in the duality formula above can be further restricted to the class of convex functions which are L_θ-Lipschitz and bounded from below. Using the fact that $W_1(\nu_n, \nu) = \sup\{\int f\, d\nu_n - \int f\, d\nu\}$ where the supremum runs over 1-Lipschitz functions and the fact that $Q_\theta\varphi$ is L_θ-Lipschitz (being an infimum of such functions), we easily get the following inequality

$$\big|\overline{\mathcal{T}}_\theta(\nu_n|\mu_n) - \overline{\mathcal{T}}_\theta(\nu|\mu)\big| \leq L_\theta W_1(\nu_n, \nu) + L_\theta W_1(\mu_n, \mu).$$

A similar (but simpler reasoning) based on the usual Kantorovich duality for $\mathcal{T}_\theta$ yields the inequality

$$\left|\mathcal{T}_\theta(\hat{\gamma}_n, \mu_n) - \mathcal{T}_\theta(\hat{\gamma}, \mu)\right| \le L_\theta W_1(\hat{\gamma}_n, \hat{\gamma}) + L_\theta W_1(\mu_n, \mu).$$

Passing to the limit as $n \to \infty$ in the identity $\overline{\mathcal{T}}_\theta(\nu_n|\mu_n) = \mathcal{T}_\theta(\hat{\gamma}_n, \mu_n)$, we end up with $\overline{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\hat{\gamma}, \mu)$.

Now it remains to extend this identity to general convex functions θ, not necessarily Lipschitz. Let $\theta : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex cost function (such that $\theta(0) = 0$) and for all $n \ge 1$, let θ_n be the convex cost function defined by $\theta_n(x) = \theta(x)$, if $x \in [0, n]$ and $\theta_n(x) = \theta(n) + \theta'(n)(x - n)$, if $x \ge n$, where θ' denotes the right derivative of θ. It is easily seen that θ_n is Lipschitz and that $Q_{\theta_n}\varphi$ converges to $Q_\theta\varphi$ monotonically as $n \to \infty$, for any function φ bounded from below. Therefore, the monotone convergence theorem implies that for any probability measure γ, it holds $\int Q_\theta\varphi\, d\gamma = \sup_{n\ge1} \int Q_{\theta_n}\varphi\, d\gamma$. We deduce from this that $\overline{\mathcal{T}}_\theta(\nu|\mu) = \sup_{n\ge1} \overline{\mathcal{T}}_{\theta_n}(\nu|\mu)$ and $\mathcal{T}_\theta(\hat{\gamma}|\mu) = \sup_{n\ge1} \mathcal{T}_{\theta_n}(\hat{\gamma}, \mu)$. Since $\overline{\mathcal{T}}_{\theta_n}(\nu|\mu) = \mathcal{T}_{\theta_n}(\hat{\gamma}, \mu)$ for all $n \ge 1$, this ends the proof of the first part of the theorem, i.e., that $\overline{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\hat{\gamma}|\mu)$.

From the first part of the theorem we conclude that there exists some $\hat{\gamma} \in \mathcal{P}_1(\mathbb{R})$ such that $\overline{\mathcal{T}}_\theta(\nu|\mu) = \mathcal{T}_\theta(\hat{\gamma}, \mu)$ for the three cost functions $\theta = \theta_1, \theta_2, \theta_1 + \theta_2$. The result then follows from the well-known additivity of $\mathcal{T}_\theta$ in dimension one: $\mathcal{T}_{\theta_1+\theta_2}(\hat{\gamma}, \mu) = \mathcal{T}_{\theta_1}(\hat{\gamma}, \mu) + \mathcal{T}_{\theta_2}(\hat{\gamma}, \mu)$. This completes the proof of the theorem. □

4. Dual formulation for weak transport-entropy inequalities

In this short section we recall the Bobkov and Götze dual formulation of the transport-entropy inequality (2) and its extensions, borrowed from [25], related to the transport-entropy inequalities of Definition 1.1, in terms of infimum convolution inequalities. The results are stated in dimension one to fit our framework but hold in more general settings (see [25]). They will be used in the next sections.

Lemma 4.1. *Let $\mu \in \mathcal{P}_1(\mathbb{R})$ and $\theta : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex cost function and, for all functions $g : \mathbb{R} \to \mathbb{R}$ bounded from below, set*

$$Q_t g(x) := \inf_{y\in\mathbb{R}} \left\{ g(y) + t\theta\left(\frac{|x-y|}{t}\right)\right\}, \quad t > 0, x \in \mathbb{R}.$$

Then the following holds.

(i) *μ satisfies $\mathrm{T}(\theta)$ if and only if for all $g : \mathbb{R} \to \mathbb{R}$ bounded from below it holds*

$$\exp\left(\int Q_1 g\, d\mu\right) \exp\left(-\int g\, d\mu\right) \le 1.$$

(ii) *μ satisfies $\overline{\mathrm{T}}^+(\theta)$ if and only if for all convex $g : \mathbb{R} \to \mathbb{R}$ bounded from below it holds*

$$\exp\left(\int Q_1 g\, d\mu\right) \int \exp(-g)\, d\mu \le 1.$$

(iii) *μ satisfies $\overline{\mathrm{T}}^-(\theta)$ if and only if for all convex $g : \mathbb{R} \to \mathbb{R}$ bounded from below it holds*

$$\int \exp(Q_1 g)\, d\mu \exp\left(-\int g\, d\mu\right) \le 1.$$

(iv) *If μ satisfies $\overline{\mathrm{T}}(\theta)$, then for all convex $g : \mathbb{R} \to \mathbb{R}$ bounded from below it holds*

$$\int \exp(Q_t g)\, d\mu \int \exp(-g)\, d\mu \le 1, \tag{12}$$

with $t = 2$. Conversely, if μ satisfies (12) for some $t > 0$, then it satisfies $\overline{\mathrm{T}}(t\theta(\cdot/t))$.

Proof. The first item is due to Bobkov and Götze [9] and is based on a combination of the well-known duality formulas for the relative entropy and for the transport cost $\mathcal{T}_\theta$. Items (ii) and (iii) generalize the first item to the framework of weak transport-entropy inequalities. We refer to [25, Proposition 4.5] for a more general statement and for a proof (based on an extension of duality for weak transport costs).

Finally we sketch the proof of item (iv) (which already appeared in a slightly different form in [22, Propositions 8.2 and 8.3]). By the definition of $\overline{\mathrm{T}}(\theta)$, if μ satisfies $\overline{\mathrm{T}}(\theta)$ then it satisfies $\overline{\mathrm{T}}^{\pm}(\theta)$ and therefore, it satisfies the exponential inequalities given in items (ii) and (iii). Note that if g is convex and bounded from below then $Q_1 g$ is also convex and bounded from below. Therefore it holds

$$\exp\left(\int Q_1 g \, d\mu\right) \int \exp(-g)\, d\mu \leq 1$$

and

$$\int \exp\bigl(Q_1(Q_1 g)\bigr)\, d\mu \exp\left(-\int Q_1 g \, d\mu\right) \leq 1.$$

Multiplying these two inequalities and noticing that $Q_1(Q_1 g) = Q_2 g$ (for a proof of this well-known semi-group property, see, e.g., [48, Theorem 22.46]) gives (12) with $t = 2$. The converse implication simply follows from Jensen's inequality. □

5. A transport form of the convex Poincaré inequality

This section is devoted to the proof of Theorem 1.6. Since, from [8], item (i) is equivalent to item (ii), and since it is easy to prove that item (iii) implies item (ii) we will mainly focus on the implication (i) $\Rightarrow$ (iii). Our strategy is to prove a modified logarithmic Sobolev inequality for the exponential probability measure τ and then, using a transport argument, a modified logarithmic Sobolev inequality for general μ (satisfying the assumption of item (i)) restricted to convex or concave Lipschitz functions. Finally, following the well-known Hamilton–Jacobi interpolation technique of [7], the desired transport inequalities will follow in their dual forms (recalled in Lemma 4.1).

We need to introduce some notations. Given a convex or concave function $g\colon \mathbb{R} \to \mathbb{R}$, we set

$$|\nabla g|(x) := \min\bigl\{\bigl|\theta g'_-(x) + (1-\theta) g'_+(x)\bigr|; \theta \in [0,1]\bigr\}, \tag{13}$$

where g'_- and g'_+ denote the left and right derivatives of g (which are well-defined and finite valued for such functions, see, e.g., [27, Theorem 6.3, Chapter 0]). In particular, if g is convex

$$|\nabla g|(x) = \begin{cases} |g'_+(x)| & \text{if } g'_+(x) \leq 0, \\ 0 & \text{if } g'_-(x) \leq 0 \leq g'_+(x), \\ g'_-(x) & \text{if } g'_-(x) \geq 0, \end{cases}$$

and if g is concave

$$|\nabla g|(x) = \begin{cases} |g'_-(x)| & \text{if } g'_-(x) \leq 0, \\ 0 & \text{if } g'_+(x) \leq 0 \leq g'_-(x), \\ g'_+(x) & \text{if } g'_+(x) \geq 0. \end{cases}$$

The following result is one of the key ingredients in the proof of Theorem 1.6. Recall that the map U_μ is defined by (4) in the Introduction.

Proposition 5.1. *Let $\mu \in \mathcal{P}_1(\mathbb{R})$. Assume that $\sup_{x\in\mathbb{R}}[U_\mu(x+1) - U_\mu(x)] \leq h$ for some $h > 0$. Set $K := 2740$ and $c := 1/(10\sqrt{2})$. Then, for all convex or concave and l-Lipschitz functions g with $l \leq c/h$, it holds*

$$\mathrm{Ent}_\mu\bigl(e^g\bigr) \leq K h^2 \int_{\mathbb{R}} \bigl|\nabla g(x)\bigr|^2 e^{g(x)} \mu(dx), \tag{14}$$

where we recall that for all non-negative functions h,

$$\mathrm{Ent}_\mu(h) = \int h \log h \, d\mu - \log\left(\int h \, d\mu\right) \int h \, d\mu.$$

The proof of Proposition 5.1 is postponed to the end of this section.

Proof of Theorem 1.6. As already mentioned above, from [8] we conclude that item (i) is equivalent to item (ii). In order to make the dependency of the constants explicit in the implication (i) $\Rightarrow$ (ii), one can use a well-known expansion argument: apply (14) to εf and take the limit $\varepsilon \to 0$, see, e.g., [5]. On the other hand, using a similar expansion argument, it is easy to prove that item (iii) implies item (ii) with $C = 2D$: apply (12) to $g = \varepsilon f$ and take the limit $\varepsilon \to 0$, see, e.g., [22,23]. Hence, we are left with the proof of (i) implies (iii) which closely follows the Hamilton–Jacobi semi-group approach introduced in [7].

Let μ be a probability measure on the line and assume that item (i) of Theorem 1.6 holds. According to Proposition 5.1, for any convex or concave differentiable function g which is l-Lipschitz with $l \le c/h := l_o$, it holds

$$\mathrm{Ent}_\mu\left(e^g\right) \le Kh^2 \int_{\mathbb{R}} \left|\nabla g(x)\right|^2 e^{g(x)} \mu(dx), \tag{15}$$

with $K = 2740$ and $c = 1/(10\sqrt{2})$. It is easy to check that the latter is equivalent to

$$\mathrm{Ent}_\mu\left(e^g\right) \le \int \alpha^*\left(|\nabla g|\right) e^g \, d\mu, \tag{16}$$

for all convex or concave $g\colon \mathbb{R} \to \mathbb{R}$ with

$$\alpha^*(v) := \sup_u \left\{uv - \alpha(u)\right\} = \begin{cases} Kh^2 v^2 & \text{if } |v| \le l_o, \\ +\infty & \text{if } |v| > l_o \end{cases}$$

the convex conjugate of

$$\alpha(u) := \begin{cases} \frac{u^2}{4Kh^2} & \text{if } |u| \le 2l_o Kh^2, \\ l_o|u| - l_o^2 Kh^2 & \text{if } |u| > 2l_o Kh^2. \end{cases}$$

Now, introduce the inf-convolution operators Q_t, for $t \in (0, 1]$, defined by

$$Q_t f(x) := \inf_{y \in \mathbb{R}} \left\{ f(y) + t\alpha\left(\frac{x-y}{t}\right)\right\}, \quad x \in \mathbb{R}, t \in (0, 1],$$

which make sense, for instance, for any Lipschitz function f or for any function f bounded from below. For simplicity denote by $\mathcal{F}$ the set of functions $f\colon \mathbb{R} \to \mathbb{R}$ that are convex and bounded below or concave and l-Lipschitz for $l \le l_o$. Then, Q_t satisfies the following properties:

(a) If f is convex, then $Q_t f$ is convex.
(b) If f is concave and Lipschitz, then $Q_t f$ is concave.
(c) If $f \in \mathcal{F}$, then $Q_t f$ is l_o-Lipschitz.
(d) If $f \in \mathcal{F}$, then the function $u(t, x) := Q_t f(x)$ satisfies the following Hamilton–Jacobi equation

$$\frac{d}{dt_+} u(t, x) + \alpha^*\left(\left|\nabla^- u\right|\right)(t, x) = 0, \quad \forall t \in (0, 1], \forall x \in \mathbb{R}, \tag{17}$$

where $\frac{d}{dt_+}$ is the right time-derivative and $|\nabla^- u(t, x)| = \limsup_{y \to x} \frac{[u(t,y) - u(t,x)]_-}{|y-x|}$ (where as usual $[X]_- := \max(-X, 0)$ denotes the negative part).

Item (a) is easy to check and is a general fact about infimum convolution of two convex functions (f and α). Item (b) follows from the fact that, after a change of variables, $Q_t f(x) = \inf_u\{f(x-u) + t\alpha(\frac{u}{t})\}$ so that $Q_t f$ is an infimum of concave functions and is therefore also concave. As for item (c), we observe that $x \mapsto t\alpha(\frac{x-y}{t})$ is l_o-Lipschitz, for any y, so that $Q_t f$ is also l_o-Lipschitz as an infimum of l_o-Lipschitz functions. A proof of item (d) can be found in [24] or [4].

With these properties and definitions in hand, let $f \in \mathcal{F}$ and (following [7]) define

$$F(t) := \frac{1}{t}\log\left(\int_{\mathbb{R}} e^{tQ_t f}\,d\mu\right), \quad t \in (0,1].$$

The function F is right differentiable at every point $t > 0$ (thanks to the above technical properties of Q_t, see, e.g., [24] for details) and it holds

$$\begin{aligned}\frac{d}{dt_+}F(t) &= \frac{1}{t^2}\frac{1}{\int_{\mathbb{R}} e^{tQ_t f}\,d\mu}\left(\mathrm{Ent}_\mu\big(e^{tQ_t f}\big) + t^2\int_{\mathbb{R}}\left(\frac{d}{dt_+}Q_t f\right)e^{tQ_t f}\,d\mu\right)\\ &= \frac{1}{t^2}\frac{1}{\int_{\mathbb{R}} e^{tQ_t f}\,d\mu}\left(\mathrm{Ent}_\mu\big(e^{tQ_t f}\big) - Kh^2t^2\int_{\mathbb{R}}\big|\nabla^- Q_t f\big|^2 e^{tQ_t f}\,d\mu\right)\\ &\le \frac{Kh^2}{\int_{\mathbb{R}} e^{tQ_t f}\,d\mu}\left(\int|\nabla Q_t f|^2 e^{tQ_t f}\,d\mu - \int_{\mathbb{R}}\big|\nabla^- Q_t f\big|^2 e^{tQ_t f}\,d\mu\right) \le 0,\end{aligned}$$

where the second equality follows from (17), the first inequality from (16) applied to the function $g = tQ_t f$ (which is convex or concave and tl_o-Lipschitz) and the last inequality from the fact that for a convex or concave function g, $|\nabla g| \le |\nabla^- g|$ (we recall that $|\nabla g|$ is defined in (13)).

Thus the function F is non-increasing and satisfies $F(1) \le \lim_{t\to 0} F(t) = \int f\,d\mu$. In other words,

$$\int e^{Q_1 f}\,d\mu \le e^{\int f\,d\mu} \quad \forall f \in \mathcal{F}. \tag{18}$$

Now according to item (iii) of Lemma 4.1 one concludes, on the one hand, that μ satisfies the transport-entropy inequality $\overline{\mathrm{T}}^-(\alpha)$: $\overline{\mathcal{T}}_\alpha(\mu|\nu) \le H(\nu|\mu)$, for all $\nu \in \mathcal{P}_1(\mathbb{R})$.

On the other hand, applying (18) to $f = -Q_1 g$ with g convex and bounded from below (so that f is concave and l_o-Lipschitz) yields to $e^{\int Q_1 g\,d\mu}\int e^{Q_1(-Q_1 g)}\,d\mu \le 1$. Since $Q_1(-Q_1 g) \ge -g$ we end up with

$$e^{\int Q_1 g\,d\mu}\int e^{-g}\,d\mu \le 1,$$

for all g convex and bounded from below. According to item (ii) of Lemma 4.1, this implies that μ satisfies the transport-entropy inequality $\overline{\mathrm{T}}^+(\alpha)$: $\overline{\mathcal{T}}_\alpha(\nu|\mu) \le H(\nu|\mu)$, for all $\nu \in \mathcal{P}_1(\mathbb{R})$, which completes the proof. □

The end of the section is dedicated to the proof of Proposition 5.1.

Proof of Proposition 5.1. Let K and c be defined by Lemma 5.2 below. We may deal first with convex functions g and divide the proof into three different (sub-)cases: g monotone (non-decreasing and then non-increasing), and g arbitrary.

Assume first that g is convex *non-decreasing* and l-Lipschitz with $l \le c/h$. Set $f = g \circ U_\mu$, with U_μ defined by (4). Then, since g is non-decreasing, and since $U_\mu(x-1) \ge U_\mu(x) - h$ by assumption, for all $x \in \mathbb{R}$, it holds

$$f(x) - f(x-1) \le g\big(U_\mu(x)\big) - g\big(U_\mu(x) - h\big) \le lh \le c, \quad \forall x \in \mathbb{R}.$$

Therefore, since μ is the image of τ under the map U_μ, Lemma 5.2 (apply (19) to f) and the latter guarantee that

$$\mathrm{Ent}_\mu\big(e^g\big) = \mathrm{Ent}_\tau\big(e^f\big) \le K\int_{\mathbb{R}}\big(f(x) - f(x-1)\big)^2 e^{f(x)}\tau(dx)$$

$$\leq K\int_{\mathbb{R}}\big(g\big(U_\mu(x)\big)-g\big(U_\mu(x)-h\big)\big)^2 e^{g(U_\mu(x))}\tau(dx)$$
$$= K\int_{\mathbb{R}}\big(g(x)-g(x-h)\big)^2 e^{g(x)}\mu(dx)\leq Kh^2\int_{\mathbb{R}}\big|\nabla g(x)\big|^2 e^{g(x)}\mu(dx),$$

where the last inequality is due to the fact that g is convex and non-decreasing and therefore satisfies $0\leq g(x)-g(x-h)\leq g'_-(x)h=|\nabla g(x)|h$. As a conclusion we proved (14) for all convex *non-decreasing* and l-Lipschitz functions g with $l\leq c/h$.

Now suppose that g is convex, *non-increasing* and l-Lipschitz with $l\leq c/h$ and set $f(x)=g(U_\mu(-x))$. The function f is non-decreasing and, since $U_\mu(-x+1)\geq U_\mu(-x)+h$ by assumption, satisfies

$$f(x)-f(x-1)=g\big(U_\mu(-x)\big)-g\big(U_\mu(-x+1)\big)\leq g\big(U_\mu(-x)\big)-g\big(U_\mu(-x)+h\big)\leq c.$$

Similarly to the previous lines, Lemma 5.2 implies that

$$\begin{aligned}\mathrm{Ent}_\mu\big(e^g\big)=\mathrm{Ent}_\tau\big(e^f\big)&\leq K\int_{\mathbb{R}}\big(g\big(U_\mu(-x)\big)-g\big(U_\mu(-x+1)\big)\big)^2 e^{g(U_\mu(-x))}\tau(dx)\\&\leq K\int_{\mathbb{R}}\big(g\big(U_\mu(-x)\big)-g\big(U_\mu(-x)+h\big)\big)^2 e^{g(U_\mu(-x))}\tau(dx)\\&= K\int_{\mathbb{R}}\big(g(x)-g(x+h)\big)^2 e^{g(x)}\mu(dx)\leq Kh^2\int_{\mathbb{R}}\big|\nabla g(x)\big|^2 e^{g(x)}\mu(dx),\end{aligned}$$

where we used the symmetry of τ and that $0\leq g(x)-g(x+h)\leq g'_+(x)(-h)=|\nabla g(x)|h$. Therefore we proved (14) for all convex *non-increasing* and l-Lipschitz functions g with $l\leq c/h$.

Finally, consider an arbitrary convex and l-Lipschitz function g with $l\leq c/h$ and assume without loss of generality that g is not monotone. Being convex, there exists some $a\in\mathbb{R}$ such that g restricted to $(-\infty,a]$ is non-increasing and g restricted to $[a,\infty)$ is non-decreasing. Subtracting $g(a)$ if necessary, one can further assume that $g(a)=0$, since (14) is invariant by the change of function $g\to g+C$ (for any constant C). Set $g_1=g\mathbf{1}_{(-\infty,a]}$ and $g_2=g\mathbf{1}_{(a,\infty)}$. The functions g_1 and g_2 are convex, monotone and l-Lipschitz. Therefore, according to the two previous sub-cases, it holds

$$\mathrm{Ent}_\mu\big(e^{g_1}\big)\leq Kh^2\int_{-\infty}^{a}\big|\nabla g(x)\big|^2 e^{g(x)}\mu(dx)\quad\text{and}\quad\mathrm{Ent}_\mu\big(e^{g_2}\big)\leq Kh^2\int_{a}^{+\infty}\big|\nabla g(x)\big|^2 e^{g(x)}\mu(dx).$$

So what remains to prove is the following sub-additivity property of the entropy functional

$$\mathrm{Ent}_\mu\big(e^{g_1+g_2}\big)\leq\mathrm{Ent}_\mu\big(e^{g_1}\big)+\mathrm{Ent}_\mu\big(e^{g_2}\big),$$

which, since $\int ge^g\,d\mu=\int g_1e^{g_1}\,d\mu+\int g_2e^{g_2}\,d\mu$, amounts to proving that

$$\int e^{g_1}\,d\mu\log\bigg(\int e^{g_1}\,d\mu\bigg)+\int e^{g_2}\,d\mu\log\bigg(\int e^{g_2}\,d\mu\bigg)\leq\int e^g\,d\mu\log\bigg(\int e^g\,d\mu\bigg).$$

Setting $A=\int e^{g_1}\,d\mu-1$, $B=\int e^{g_2}\,d\mu-1$ and $X=\int e^g\,d\mu$ and observing that $A+B+1=X$ the latter is equivalent to proving that

$$(A+1)\log(A+1)+(B+1)\log(B+1)\leq X\log X,$$

which follows from the sub-additivity property of the convex function $\Phi\colon x\mapsto(x+1)\log(x+1)$ on $[0,\infty)$, that satisfies $\Phi(0)=0$. This completes the proof when g is convex.

The case g concave follows the same lines (use (20) instead of (19)). Details are left to the reader. □

In the proof of Proposition 5.1, we used the following lemma which is a (discrete) variant of a result by Bobkov and Ledoux [11] and an entropic counterpart of a result (involving the variance) by Bobkov and Götze (see [8, Lemma 4.8]).

Lemma 5.2. *For all non-decreasing functions $f:\mathbb{R}\to\mathbb{R}$ with $f(x)-f(x-1)\le 1/(10\sqrt{2})$, $x\in\mathbb{R}$, it holds*

$$\mathrm{Ent}_\tau\big(e^f\big)\le 939\int_{\mathbb{R}}\big(f(x)-f(x-1)\big)^2e^f\,d\tau \tag{19}$$

and

$$\mathrm{Ent}_\tau\big(e^f\big)\le 2740\int_{\mathbb{R}}\big(f(x+1)-f(x)\big)^2e^f\,d\tau. \tag{20}$$

Proof. Let τ^+ be the exponential probability measure on $\mathbb{R}^+$: $\tau^+(dx)=e^{-x}\mathbb{1}_{[0,\infty)}\,dx$. We shall use the following fact, borrowed from [8, Lemma 4.7] (with $a=0$ and $h=1$ so that the constant $c(a,h)$ appearing in [8] can be explicitly bounded by $1/200$): for all $f:[-1,\infty)\to\mathbb{R}$ non-decreasing and satisfying $f(0)=0$, it holds

$$\int f^2\,d\tau^+\le 200\int\big(f(x)-f(x-1)\big)^2\,d\tau^+(x). \tag{21}$$

We will first prove (19). Since (19) is invariant by the change of function $f\to f+C$ for any constant C, we may assume without loss of generality that $f(0)=0$. Set $\tilde f(y):=-f(-y)$, $y\in\mathbb{R}$ and observe that f is non-decreasing. Since $u\log u\ge u-1$ for all $u\ge 0$, one has

$$\begin{aligned}\mathrm{Ent}_\tau\big(e^f\big)&\le\int\big(fe^f-e^f+1\big)\,d\tau\\&=\int\left(\int_0^1 tf^2e^{tf}\,dt\right)d\tau\\&=\frac12\int_0^\infty f^2\left(\int_0^1 te^{tf}\,dt\right)d\tau^+ +\frac12\int_0^\infty\tilde f^2\left(\int_0^1 te^{-t\tilde f}\,dt\right)d\tau^+\\&\le\frac14\int f^2e^f\,d\tau^+ +\frac14\int\tilde f^2\,d\tau^+,\end{aligned} \tag{22}$$

where the last inequality comes from the fact both f and $\tilde f$ are non-negative on $\mathbb{R}^+$. Now suppose that the function f is such that $f(y)-f(y-1)\le c$ for all $y\in\mathbb{R}$ and some $c\in(0,1)$. Our aim is to bound each term in the right hand side of the above inequality.

By (21) applied to the function $\tilde f$, one has

$$\begin{aligned}\int\tilde f^2\,d\tau^+&\le 200\int\big(\tilde f(y)-\tilde f(y-1)\big)^2\,d\tau^+(y)\\&=\frac{200}{\int e^{-\tilde f}\,d\nu}\int\big(\tilde f(y)-\tilde f(y-1)\big)^2e^{-\tilde f(y)}\,d\tau^+(y)\\&\le 200\exp\left(\frac{\int\tilde f(y)(\tilde f(y)-\tilde f(y-1))^2\,d\tau^+(y)}{\int(\tilde f(y)-\tilde f(y-1))^2\,d\tau^+(y)}\right)\int\big(\tilde f(y)-\tilde f(y-1)\big)^2e^{-\tilde f(y)}\,d\tau^+(y),\end{aligned}$$

where we set $\frac{d\nu}{d\tau^+}(y)=\frac{(\tilde f(y)-\tilde f(y-1))^2}{\int(\tilde f(y)-\tilde f(y-1))^2\,d\tau^+(y)}$ and we used Jensen's inequality to guarantee that $1/\int e^{-\tilde f}\,d\nu\le e^{\int\tilde f\,d\nu}$. By the Cauchy–Schwarz inequality and using (21) again, we get

$$\begin{aligned}\int\tilde f(y)\big(\tilde f(y)-\tilde f(y-1)\big)^2\,d\tau^+(y)&\le\left(\int\big(\tilde f(y)-\tilde f(y-1)\big)^4\,d\tau^+(y)\right)^{1/2}\left(\int\tilde f^2\,d\tau^+\right)^{1/2}\\&\le\sqrt{200}c\int\big(\tilde f(y)-\tilde f(y-1)\big)^2\,d\tau^+(y).\end{aligned}$$

It finally follows that

$$\begin{aligned}\int \tilde f^2\,d\tau^+ &\le 200e^{\sqrt{200}c}\int \big(\tilde f(y)-\tilde f(y-1)\big)^2 e^{-\tilde f(y)}\,d\tau^+(y)\\ &= 200e^{\sqrt{200}c}\int_{-\infty}^{0}\big(f(y+1)-f(y)\big)^2 e^{f(y)}e^{y}\,dy\\ &= 200e^{\sqrt{200}c-1}\int_{-\infty}^{1}\big(f(y)-f(y-1)\big)^2 e^{f(y-1)}e^{y}\,dy\\ &\le 400e^{\sqrt{200}c+1}\int_{-\infty}^{1}\big(f(y)-f(y-1)\big)^2 e^{f(y)}\,d\tau(y),\end{aligned}$$

where in the last line we used that $e^y/(e^{-|y|}/2)\le 2e^2$ for all $y\le 1$.

Next we deal with the first term in the right hand side of (22). Our aim is to apply (21) to $g=fe^{f/2}$. Observe that, since f is non-decreasing, $f(x)\ge f(-1)\ge -c+f(0)=-c\ge -1$ so that, since $x\mapsto xe^{x/2}$ is non-increasing on $[-2,\infty)$, we are guaranteed that g is non-decreasing on $[-1,\infty)$. Therefore we can apply (21) to g. Applying (21) to $g=fe^{f/2}$ and using the inequality

$$0\le be^{b/2}-ae^{a/2}\le (b-a)e^{b/2}+\frac{b}{2}(b-a)e^{b/2},\quad -2\le a\le b,$$

we get

$$\begin{aligned}B:=\int f^2e^f\,d\tau^+ &\le 200\int\big(f(y)e^{f(y)/2}-f(y-1)e^{f(y-1)/2}\big)^2\,d\tau^+(y)\\ &\le 400\int\big(f(y)-f(y-1)\big)^2e^{f(y)}\,d\tau^+(y)+100\int f^2(y)\big(f(y)-f(y-1)\big)^2e^{f(y)}\,d\tau^+(y)\\ &\le 400\int\big(f(y)-f(y-1)\big)^2e^{f(y)}\,d\tau^+(y)+100c^2B.\end{aligned}$$

Therefore, provided $c<1/10$, we end up with

$$B\le 400/\big(1-100c^2\big)\int\big(f(y)-f(y-1)\big)^2e^{f(y)}\,d\tau^+(y).$$

Hence, plugging the previous two bounds into (22) and choosing $c=1/\sqrt{200}$, Inequality (19) follows with the constant 939.

To obtain (20) from (19), it suffices to observe that, by a simple change of variable

$$\begin{aligned}\int\big(f(y)-f(y-1)\big)^2e^{f(y)}\,d\tau(y) &= \int\big(f(x+1)-f(x)\big)^2e^{f(x+1)}\frac{e^{-|y+1|}}{2}\,dy\\ &\le e^{c+1}\int\big(f(x+1)-f(x)\big)^2e^{f(x)}\,d\tau(x)\end{aligned}$$

and that $939e^{c+1}\le 2740$ for $c=1/\sqrt{200}$. This ends the proof. □

6. Proof of Theorem 1.3

In this final section we prove Theorem 1.3. As mentioned in the introduction, we may need to decompose the cost θ into the sum of two costs, one that takes care of the behavior near 0 (the cost θ_1) and the other one vanishing in a neighborhood of 0 (the cost θ_2). The next theorem deals with the latter.

Theorem 6.1. *Let $\mu \in \mathcal{P}_1(\mathbb{R})$ and $\beta : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex cost function such that $\{t \in \mathbb{R}^+ : \beta(t) = 0\} = [0, t_o]$, where $t_o > 0$ is some positive constant. The following are equivalent*:

(i) *There exists $a > 0$ such that μ satisfies the transport-entropy inequality $\mathrm{T}(\beta(a\cdot))$.*
(ii) *There exists $a' > 0$ such that μ satisfies the weak transport-entropy inequality $\overline{\mathrm{T}}(\beta(a'\cdot))$.*
(iii) *There exists $b > 0$ such that $\max(K^+(b), K^-(b)) < \infty$, where*

$$K^+(b) := \sup_{x \geq m} \frac{1}{\mu((x,\infty))} \int_x^\infty e^{\beta(b(u-x))} \mu(du),$$

$$K^-(b) := \sup_{x \leq m} \frac{1}{\mu((-\infty,x))} \int_{-\infty}^x e^{\beta(b(x-u))} \mu(du),$$

and m is a median of μ. (Here we use the convention $0/0 = 0$.)
(iv) *There exists $d > 0$ such that*

$$\big|U_\mu(u) - U_\mu(v)\big| \leq \frac{1}{d} \beta^{-1}\big(|u-v|\big), \quad \forall u \neq v \in \mathbb{R}.$$

(*Note that β^{-1} is well defined on $(0,\infty)$.*)

In particular,

- (i) $\Rightarrow$ (ii) *with* $a' = a$,
- (ii) $\Rightarrow$ (iii) *with* $b = a'/2$,
- (ii) $\Rightarrow$ (iv) *with* $d = a' \frac{t_o}{8\beta^{-1}(\log 3)}$,
- (iv) $\Rightarrow$ (ii) *with* $a' = d \frac{t_o}{9\beta^{-1}(2)}$.

Proof of Theorem 6.1. The equivalence between items (i), (iii) and (iv) is proved in [21, Theorem 2.2], with some explicit dependency between the constants. In order to complete the proof of Theorem 6.1 we need to show that (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii).

It follows from Jensen's inequality that $\mathcal{T}_{\beta(a\cdot)}(\mu,\nu) \geq \max(\overline{\mathcal{T}}_{\beta(a\cdot)}(\nu|\mu); \overline{\mathcal{T}}_{\beta(a\cdot)}(\mu|\nu))$. Therefore (i) implies (ii) with $a' = a$.

Next we will show that (ii) implies (iii). Assume that μ satisfies $\overline{\mathrm{T}}(\beta(a'\cdot))$ for some $a' > 0$. According to item (iv) of Lemma 4.1, for all convex functions $g : \mathbb{R} \to \mathbb{R}$ bounded from below, it holds

$$\int \exp(Qf)\, d\mu \int e^{-f}\, d\mu \leq 1, \quad \text{where } Qf(x) = \inf_{y \in \mathbb{R}} \big\{ f(y) + 2\beta\big(a'|y-x|/2\big) \big\}.$$

Consider the convex function f_x which equals 0 on $(-\infty, x]$ and ∞ otherwise. Then $Qf_x(y) = 0$ on $(-\infty, x]$ and $Qf_x(y) = 2\beta(a'(y-x)/2)$ on (x,∞). Thus, applying the inequality above to f_x yields

$$\Big(\mu\big((-\infty,x]\big) + \int_{(x,\infty)} e^{2\beta(a'(y-x)/2)} \mu(dy)\Big) \mu\big((-\infty,x]\big) \leq 1.$$

Considering $x \geq m$ yields that $K^+(a'/2) \leq 3$. One proves similarly that $K^-(a'/2) \leq 3$. This shows that (ii) implies (iii) with $b = a'/2$. This achieves the proof of Theorem 6.1. □

We are now in a position to prove Theorem 1.3.

Proof of Theorem 1.3. Let $\theta : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex cost function such that $\theta(t) = t^2$ on $[0, t_o]$ for some $t_o > 0$. Let us define $\theta_1(t) = t^2$ on $[0, t_o]$ and $\theta_1(t) = 2tt_o - t_o^2$ on $[t_o, +\infty)$ and $\theta_2(t) = [\theta(t) - t_o^2]_+$. Note that θ_1 and θ_2 are both convex and that θ_2 vanishes on $[0, t_o]$ and that $\max(\theta_1, \theta_2) \leq \theta \leq \theta_1 + \theta_2$.

First assume that μ satisfies the weak transport-entropy inequality $\overline{\mathrm{T}}(\theta(a\cdot))$ for some $a > 0$ (i.e., item (i) of Theorem 1.3). Then, since $\theta \geq \theta_2$ it clearly satisfies $\overline{\mathrm{T}}(\theta_2(a\cdot))$. According to Theorem 6.1, the mapping U_μ sending the exponential measure on μ satisfies the condition:

$$\sup_{x\in\mathbb{R}} U_\mu(x+u) - U_\mu(x) \leq \frac{1}{b}\theta_2^{-1}(u), \quad \forall u > 0, \tag{23}$$

with $b = a\kappa_1$, where $\kappa_1 = t_o/(8\theta_2^{-1}(\log 3))$. Since $\theta_2^{-1}(u) = \theta^{-1}(u + t_o^2)$, this proves item (ii) of Theorem 1.3 with the announced dependency between the constants.

Now assume that μ satisfies item (ii) of Theorem 1.3, or equivalently (23) for some $b > 0$. Recall that we set, in Theorem 1.6, $\kappa := 5480$ and $c := 1/(10\sqrt{2})$. Then, observe that, plugging $u = 1$ into (23) and using Theorem 1.6, one concludes that μ satisfies $\overline{\mathrm{T}}(\alpha)$ with α defined by $\alpha(u) = \bar{\alpha}(u/\sqrt{2D})$, with $D = \kappa(\theta^{-1}(1+t_o^2))^2\frac{1}{b^2}$ and

$$\bar{\alpha}(v) = \begin{cases} v^2 & \text{if } |v| \leq c\sqrt{\kappa/2}, \\ c\sqrt{\kappa}|v| - \frac{c^2\kappa}{2} & \text{if } |v| > c\sqrt{\kappa/2} \end{cases} = \begin{cases} v^2 & \text{if } |v| \leq \sqrt{137/10}, \\ 2\sqrt{137}|v| - \frac{137}{5} & \text{if } |v| > c\sqrt{137/10}. \end{cases}$$

It is not difficult to check that $\bar{\alpha}$ compares to θ_1. More precisely, for all $v \in \mathbb{R}$, it holds

$$\bar{\alpha}(v) \geq \theta_1\left(\max\left(\frac{c\sqrt{\kappa/2}}{t_o}; 1\right)|v|\right) = \theta_1\left(\max\left(\frac{\sqrt{137/10}}{t_o}; 1\right)|v|\right).$$

Therefore μ satisfies $\overline{\mathrm{T}}(\theta_1(a_1'\cdot))$, and by monotonicity $\overline{\mathrm{T}}(\theta_1(a_1\cdot))$ with

$$a_1' := \frac{\max((c\sqrt{\kappa/2})/t_o; 1)}{\sqrt{2\kappa}\theta^{-1}(1+t_o^2)}b = \frac{\max(\frac{\sqrt{137/10}}{t_o}; 1)}{4\sqrt{685}\theta^{-1}(1+t_o^2)}b \geq \frac{1}{105}\frac{\max(1,t_o)}{t_o\theta^{-1}(1+t_o^2)}b =: a_1.$$

On the other hand, according to Theorem 6.1, μ also satisfies $\overline{\mathrm{T}}(\theta_2(a_2\cdot))$, with $a_2 = \frac{t_o}{\theta^{-1}(2+t_o^2)}b$. Letting $a = \min(a_1, a_2)$, one concludes that μ satisfies both $\overline{\mathrm{T}}(\theta_1(a\cdot))$ and $\overline{\mathrm{T}}(\theta_2(a\cdot))$. Hence, since $\theta(at) \leq \theta_1(at) + \theta_2(at)$ and according to (7), it holds

$$\begin{aligned}\mathcal{T}_{\theta(a\cdot)}(\nu|\mu) &\leq \mathcal{T}_{\theta_1(a\cdot)+\theta_2(a\cdot)}(\nu|\mu) = \mathcal{T}_{\theta_1(a\cdot)}(\nu|\mu) + \mathcal{T}_{\theta_2(a\cdot)}(\nu|\mu) \\ &\leq 2H(\nu|\mu),\end{aligned}$$

and so μ satisfies $\overline{\mathrm{T}}^+(\frac{1}{2}\theta(a\cdot))$. By convexity of θ and since $\theta(0) = 0$, it holds $\frac{1}{2}\theta(2at) \geq \theta(at)$, and so μ satisfies $\overline{\mathrm{T}}^+(\theta((a/2)\cdot))$. Finally we observe that

$$\frac{a}{2} = \frac{b}{210}\min\left(\frac{\max(1,t_o)}{t_o\theta^{-1}(1+t_o^2)}; \frac{105t_o}{\theta^{-1}(2+t_o^2)}\right) \geq \frac{b}{210}\frac{\min(1,t_o)}{\theta^{-1}(2+t_o^2)} =: \kappa_2 b$$

so that, by monotonicity, μ satisfies $\overline{\mathrm{T}}^+(\theta(\kappa_2 b\cdot))$. The same reasoning yields the conclusion that μ satisfies $\overline{\mathrm{T}}^-(\theta(\kappa_2 b\cdot))$, which completes the proof. □

Acknowledgements

The authors would like to warmly thank Greg Blekherman for discussions on the geometric aspects in Section 2.3, including his help with the proof of Theorem 2.10. They also want to thank an anonymous referee for a thorough reading of the manuscript and a number of helpful suggestions.

The authors acknowledge the kind support of the American Institute of Mathematics (AIM).

Prasad Tetali is grateful for the hospitality of Université Paris Est Marne La Vallée.

References

[1] R. Adamczak and M. Strzelecki. On the convex Poincaré inequality and weak transportation inequalities. *Bernoulli*. To appear, 2018. Available at arXiv:1703.01765v2.

[2] R. Adamczak and M. Strzelecki. Modified log-Sobolev inequalities for convex functions on the real line. Sufficient conditions. *Studia Math.* **230** (1) (2015) 59–93. MR3456588

[3] R. Adamczak. Logarithmic Sobolev inequalities and concentration of measure for convex functions and polynomial chaoses. *Bull. Pol. Acad. Sci. Math.* **53** (2) (2005) 221–238. MR2163396

[4] L. Ambrosio, N. Gigli and G. Savaré. Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below. *Invent. Math.* **195** (2) (2014) 289–391. MR3152751

[5] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques, Panoramas et Synthèses [Panoramas and Syntheses]*, **10**. Société Mathématique de France, Paris, 2000. With a preface by Dominique Bakry and Michel Ledoux. MR1845806

[6] A. Björner, M. Las Vergnas, B. Sturmfels, N. White and G. M. Ziegler. *Oriented matroids*, 2nd edition. *Encyclopedia of Mathematics and its Applications* **46**. Cambridge University Press, Cambridge, 1999. MR1744046

[7] S. G. Bobkov, I. Gentil and M. Ledoux. Hypercontractivity of Hamilton–Jacobi equations. *J. Math. Pures Appl. (9)* **80** (7) (2001) 669–696. MR1846020

[8] S. G. Bobkov and F. Götze. Discrete isoperimetric and Poincaré-type inequalities. *Probab. Theory Related Fields* **114** (2) (1999) 245–277. MR1701522

[9] S. G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** (1) (1999) 1–28. MR1682772

[10] S. G. Bobkov and C. Houdré. Weak dimension-free concentration of measure. *Bernoulli* **6** (4) (2000) 621–632. MR1777687

[11] S. G. Bobkov and M. Ledoux. Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution. *Probab. Theory Related Fields* **107** (3) (1997) 383–400. MR1440138

[12] V. I. Bogachev. *Measure theory. Vol. I*. Springer-Verlag, Berlin, 2007. MR2267655

[13] S. Boucheron, G. Lugosi and P. Massart. *Concentration inequalities. A nonasymptotic theory of independence*. Oxford University Press, Oxford, 2013. With a foreword by Michel Ledoux. MR3185193

[14] S. Cambanis, G. Simons and W. Stout. Inequalities for $Ek(X, Y)$ when the marginals are fixed. *Z. Wahrsch. Verw. Gebiete* **36** (4) (1976) 285–294. MR0420778

[15] G. Dall'Aglio. Sugli estremi dei momenti delle funzioni di ripartizione doppia. *Ann. Sc. Norm. Super. Pisa (3)* **10** (1956) 35–74. MR0081577

[16] H. Djellout, A. Guillin and L. Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.* **32** (2004) 2702–2732. MR2078555

[17] R. M. Dudley. *Real analysis and probability*. *Cambridge Studies in Advanced Mathematics* **74**. Cambridge University Press, Cambridge, 2002. MR1932358

[18] N. Feldheim, A. Marsiglietti, P. Nayar and J. Wang. A note on the convex infimum convolution inequality. *Bernoulli*. To appear, 2018. Available at arXiv:1505.00240.

[19] M. Fréchet. Sur les tableaux dont les marges et des bornes sont données. *Rev. Inst. Int. Stat.* **28** (1960) 10–32. MR0115232

[20] N. Gozlan. A characterization of dimension-free concentration in terms of transportation inequalities. *Ann. Probab.* **37** (6) (2009) 2480–2498. MR2573565

[21] N. Gozlan. Transport-entropy inequalities on the line. *Electron. J. Probab.* **17** (49) (2012) 18. MR2946156

[22] N. Gozlan and C. Léonard. Transport inequalities. A survey. *Markov Process. Related Fields* **16** (4) (2010) 635–736. MR2895086

[23] N. Gozlan, C. Roberto and P.-M. Samson. A new characterization of Talagrand's transport-entropy inequalities and applications. *Ann. Probab.* **39** (3) (2011) 857–880. MR2789577

[24] N. Gozlan, C. Roberto and P.-M. Samson. Hamilton Jacobi equations on metric spaces and transport entropy inequalities. *Rev. Mat. Iberoam.* **30** (1) (2014) 133–163. MR3186934

[25] N. Gozlan, C. Roberto, P.-M. Samson and P. Tetali. Kantorovich duality for general transport costs and applications. *J. Funct. Anal.* **272** (11) (2017) 3327–3405. MR3706606

[26] G. H. Hardy, J. E. Littlewood and G. Pólya. Some simple inequalities satisfied by convex function. *Messenger Math.* **58** (1929) 145–152. MR0083530

[27] J. B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of convex analysis*. *Grundlehren Text Editions*. Springer-Verlag, Berlin, 2001. MR1865628

[28] F. Hirsch, C. Profeta, B. Roynette and M. Yor. *Peacocks and associated martingales, with explicit constructions*. *Bocconi & Springer Series* **3**. Springer, Milan; Bocconi University Press, Milan, 2011. MR2808243

[29] W. Hoeffding. Maßstabinvariante korrelationstheorie. *Schr. Math. Inst. Inst. Angew. Math. Univ. Berlin* **5** (1940) 181–233. MR0004426

[30] W. Johnson and G. Schechtman. Remarks on Talagrand's deviation inequality for Rademacher functions. Functional analysis. In *Lecture Notes in Math.* 72–77. *Austin, TX, 1987/1989*. *Longhorn Notes* **1470**. Springer, Berlin, 1991. MR1126739

[31] M. Ledoux. *The concentration of measure phenomenon*. *Mathematical Surveys and Monographs* **89**. American Mathematical Society, Providence, RI, 2001. MR1849347

[32] J. Lott and C. Villani. Hamilton-Jacobi semigroup on length spaces and applications. *J. Math. Pures Appl. (9)* **88** (3) (2007) 219–229. MR2355455

[33] A. W. Marshall, I. Olkin and B. C. Arnold. *Inequalities: theory of majorization and its applications*, 2nd edition. *Springer Series in Statistics*. Springer, New York, 2011. MR2759813

[34] K. Marton. A simple proof of the blowing-up lemma. *IEEE Trans. Inform. Theory* **32** (3) (1986) 445–446. MR0838213
[35] K. Marton. Bounding $\bar{d}$-distance by informational divergence: A method to prove measure concentration. *Ann. Probab.* **24** (2) (1996) 857–866. MR1404531
[36] K. Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.* **6** (3) (1996) 556–571. MR1392329
[37] K. Marton. An inequality for relative entropy and logarithmic Sobolev inequalities in Euclidean spaces. *J. Funct. Anal.* **264** (1) (2013) 34–61. MR2995699
[38] B. Maurey. Some deviation inequalities. *Geom. Funct. Anal.* **1** (2) (1991) 188–197. MR1097258
[39] R. Rado. An inequality. *J. Lond. Math. Soc. (2)* **27** (1952) 1–6. MR0045168
[40] P.-M. Samson. Concentration inequalities for convex functions on product spaces. In *Stochastic inequalities and applications* 33–52. *Progr. Probab.* **56**. Birkhäuser, Basel, 2003. MR2073425
[41] Y. Shu. From Hopf–Lax formula to optimal weak transfer plan. ArXiv preprint, 2016. Available at arXiv:1609.03405v1.
[42] Y. Shu and M. Strzelecki. A characterization of a class of convex log-Sobolev inequalities on the real line. *Ann. Inst. Henri Poincaré B, Probab. Stat.* To appear, 2018. Available at arXiv:1702.04698v1.
[43] S. Volker. The existence of probability measures with given marginals. *Ann. Math. Stat.* **36** (1965) 423–439. MR0177430
[44] M. Strzelecka, M. Strzelecki and T. Tkocz. On the convex infimum convolution inequality with optimal cost function. ArXiv preprint, 2017. Available at arXiv:1702.07321v1.
[45] M. Talagrand. An isoperimetric theorem on the cube and the Kintchine–Kahane inequalities. *Proc. Amer. Math. Soc.* **104** (3) (1988) 905–909. MR0964871
[46] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. Inst. Hautes Études Sci.* **81** (1995) 73–205. MR1361756
[47] M. Talagrand. Transportation cost for Gaussian and other product measures. *Geom. Funct. Anal.* **6** (3) (1996) 587–600. MR1392331
[48] C. Villani. *Optimal transport: Old and new*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer-Verlag, Berlin, 2009. MR2459454
[49] N.-Y. Wang. Concentration inequalities for Gibbs sampling under d_{l_2}-metric. *Electron. Commun. Probab.* **19** (63) (2014) 11. MR3262069
[50] L. Wu. Poincaré and transportation inequalities for Gibbs measures under the Dobrushin uniqueness condition. *Ann. Probab.* **34** (5) (2006) 1960–1989. MR2271488