CHARACTERIZATION OF A CLASS OF WEAK TRANSPORT-ENTROPY INEQUALITIES ON THE LINE

NATHAEL GOZLAN, CYRIL ROBERTO, PAUL-MARIE SAMSON, YAN SHU, PRASAD TETALI

Abstract. We study an optimal 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 in dimension one. In particular, we obtain a weak transport-entropy form of the convex Poincaré inequality in dimension one.

1. Introduction

In all the paper, $\mathcal{P}(\mathbb{R})$ denotes the set of Borel probability measures in $\mathbb{R}$ and $\mathcal{P}_1(\mathbb{R})$ the subset of probability measures having a finite first moment.

Let $\theta : \mathbb{R}^+ \to \mathbb{R}^+$ be a measurable function; the usual optimal transport cost in the sense of Kantorovich between two probability measures μ and ν on $\mathbb{R}$ is defined by

$$\mathcal{T}_\theta(\nu, \mu) = \inf_\pi \iint \theta(|x-y|)\, \pi(dxdy),$$

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 [23, 24, 25] and Talagrand [31], 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

$$\mathcal{T}_\theta(\nu, \mu) \leqslant \mathrm{H}(\nu|\mu), \qquad \forall \nu \in \mathcal{P}(\mathbb{R}), \tag{1}$$

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. Here we focus on the one dimensional case, but these definitions easily generalizes to probability measures on general metric space. As a special case, if Inequality (1) holds for a cost function of the form $\theta(x) = x^2/C$ for some $C > 0$, one says that μ satisfies the inequality $\mathrm{T}_2(C)$ (also often referred to as "Talagrand's inequality" in the literature). This inequality is for instance satisfied with constant $C = 2$ by the standard Gaussian measure as proved in [31]. We refer to the books or survey [21, 16, 32, 8] for a complete presentation of transport-entropy inequalities as well as for bibliographic references on the field, but let

Date: September 14, 2015.

1991 *Mathematics Subject Classification.* 60E15, 32F32 and 26D10.

Key words and phrases. Transport inequalities, concentration of measure, majorization.

Supported by the grants ANR 2011 BS01 007 01, ANR 10 LABX-58; the last author is supported by the NSF grants DMS 1101447 and 1407657, and is also grateful for the hospitality of Université Paris Est Marne La Vallée. The authors acknowledge the kind support of the American Institute of Mathematics (AIM).

us shortly discuss in the next few lines the consequences in terms of concentration of this type of inequalities in the special case of the inequality T_2.

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

$$\mu^n(f > \mathrm{med}(f) + t) \leqslant e^{-(t-t_o)^2/C}, \qquad \forall t \geqslant t_o := \sqrt{C\log(2)}, \tag{2}$$

where $\mathrm{med}(f)$ denotes the median of f under μ^n. We refer to [21, 8] 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 [14] that a probability μ satisfying (2) necessarily satisfies $\mathrm{T}_2(C)$, thus giving to this inequality T_2 a special status among other functional inequalities appering in the concentration of measure literature. The key argument explaining why Talagrand's inequality implies this dimension-free concentration behavior, is a well known tensorisation property enjoyed by inequalities of the form $\mathrm{T}(\theta)$ (explained in full generality in [16]) that shows in particular that if μ satisfies $\mathrm{T}_2(C)$, then the product measure $\mu \otimes \cdots \otimes \mu$ also satisfies T_2 (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 [11, 26, 33]). Then, the transport-entropy inequality for the whole measure leads again to concentration properties using the same classical arguments as in the product case. The problem is thus reduced to verify one dimensional transport-entropy inequalities and therefore, it is of a real interest to characterize the probability measures μ on $\mathbb{R}$ satisfying $\mathrm{T}(\theta)$ for a general cost function θ.

In this direction, the first named author has obtained in [15] necessary and sufficient conditions for the transport-entropy $\mathrm{T}(\theta)$ to hold, when the cost function $\theta : \mathbb{R}^+ \to \mathbb{R}^+$ is continuous, convex and quadratic near 0. Let F_μ denote the cumulative distribution function of a probability measure μ defined by

$$F_\mu(x) := \mu(-\infty, x], \qquad \forall x \in \mathbb{R},$$

and F_μ^{-1} denote its general inverse defined by

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

With these notations, the conditions obtained in [15] are expressed in terms of the behavior of the modulus of continuity of the non-decreasing map U_μ defined by

$$U_\mu := F_\mu^{-1} \circ F_\tau,$$

where τ is the symmetric exponential distribution on $\mathbb{R}$:

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

This map, which can also be expressed as follows

$$U_\mu(x) = \begin{cases} F_\mu^{-1}\left(1 - \frac{1}{2}e^{-|x|}\right) & \text{if } x \geqslant 0 \\ F_\mu^{-1}\left(e^{-|x|}\right) & \text{if } x \leqslant 0 \end{cases}$$

is the unique left-continuous and non-decreasing map transporting τ on μ. In the special case of the inequality T_2, the characterization of [15] reads as follows: a probability measure μ satisfies $\mathrm{T}_2(C)$ for some C if and only if U_μ satisfies the condition

$$\sup_{x\in\mathbb{R}} (U_\mu(x+u) - U_\mu(x)) \leqslant \frac{1}{b}\sqrt{1+u}, \qquad \forall u \geqslant 0$$

for some constant $b > 0$ and μ satisfies Poincaré inequality

$$\mathrm{Var}_\mu(f) \leqslant c \int (f')^2 \, d\mu, \tag{3}$$

for some constant $c > 0$, for all function f of class $\mathcal{C}^1$ on $\mathbb{R}$. We refer to [15] for a precise quantitative relation between C, b, c.

In the present paper, partly following [15], we focus on the study of a new weak transport-entropy inequality introduced in [18] that is related to a weaker 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 $\pi(dxdy) = p(x, dy)\mu(dx)$ of μ and ν, and where $p(x, \cdot\,)$ denotes the disintegration kernel of π with respect to its first marginal. Note that, in terms of random variables, one has the following interpretation

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

whereas

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

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 inequality, one has

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

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 transport-entropy inequality

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

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

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

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

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

In Section 4, we recall 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 first introduced by Maurey [27] and developed in [29].

These weak transport-entropy inequalities are of particular interest since the class of measures satisfying such inequalities also includes discrete measures on $\mathbb{R}$, for examples, Bernoulli, binomial and Poisson measures [18, 29]. In comparison, the classical Talagrand's transport inequality is never satisfied by a discrete probability measure (unless it is a Dirac). Indeed, as mentioned above, the Poincaré inequality is a consequence of Talagrand's transport inequality that forces the support of μ to be connected. Moreover these weak transport-entropy inequalities also enjoy a nice tensorisation property (see [18, Theorem 4.11]) that connects them to a special dimension-free concentration behavior. For instance, as shown in [18, Corollary 5.11], 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, for all positive integer n,

$$\mu^n(f > \mathrm{med}(f) + t) \leqslant e^{-(t-t_o)^2/C'}, \qquad \forall t \geqslant t_o,$$

for all *convex* and all *concave* function $f : \mathbb{R}^n \to \mathbb{R}$ which is 1-Lipsvchitz for the Euclidean norm on $\mathbb{R}^n$, where $t_o, C' > 0$ are constants related to C (see [18] for a precise and more general statement).

The main result of the paper is the following characterization of the transport inequalities $\overline{\mathrm{T}}(\theta)$ associated to convex cost functions θ which are quadratic near 0.

Theorem 1.2. *Let $\mu \in \mathcal{P}_1(\mathbb{R})$ and $\theta : \mathbb{R}^+ \to \mathbb{R}^+$ be a convex cost function such that $\theta(t) = t^2$ for all $t \leqslant t_o$, for some $t_o > 0$. The following propositions are equivalent:*

i) There exists $a > 0$ such that μ satisfies $\overline{\mathrm{T}}(\theta(a\,\cdot))$.

ii) There exists b such that for all $u > 0$,

$$\sup_x \left(U_\mu(x+u) - U_\mu(x)\right) \leqslant \frac{1}{b}\theta^{-1}(u + t_o^2).$$

Moreover, constants are related as follows : i) implies ii) with $b = a\kappa_1$ and ii) implies i), with $a = b\kappa_2$, where κ_1 and κ_2 are two constants depending only on θ. More precisely,

$$\kappa_1 = \frac{t_o}{8\theta^{-1}(\log(3) + t_o^2)},$$

and

$$\kappa_2 = \frac{1}{2}\min\left(\frac{t_o}{\theta^{-1}(2+t_o^2)}; \frac{\max((c\sqrt{K})/t_o; 1)}{2\sqrt{K}\theta^{-1}(1+t_o^2)}\right).$$

In comparison with the characterization of the inequalities $\mathrm{T}(\theta)$ given in [15], one sees that only the condition on the modulus of continuity of U_μ remains. Nevertheless, as we shall explain below, Poincaré inequality has not completely disapeared from the picture. Denoting by Δ_μ the modulus of continuity of U_μ defined by

$$\Delta_\mu(h) = \sup\left\{U_\mu(x+u) - U_\mu(x), x \in \mathbb{R}, 0 \leqslant u \leqslant h\right\}, \qquad h \geqslant 0,$$

the condition *ii)* asserts that

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

Therefore Δ_μ is bounded around zero but does not necessarily go to zero as h goes to zero. Actually if the measure μ is discrete and not a Dirac measure, the support of μ is not connected and 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 $h > 0$,

$$b - a \leqslant \Delta_\mu(h).$$

This shows that in a discrete setting $\lim_{h\to 0}\Delta_\mu(h) > 0$.

The proof of Theorem 1.2, given in Section 6, is based on some new results of independent interest. To introduce them and for a better understanding of the paper, let us briefly give the main ideas of this proof. As in the paper [15], the weak transport-entropy inequality *i)* follows from condition *ii)* by decomposition of the weak optimal cost into two parts. One part is related to the quadratic behavior of θ on $[0, t_o]$, the other part is related to its behavior for $t \geqslant t_o$: one has $\theta \leqslant \theta_1 + \theta_2$ with

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

and

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

Therefore,

$$\mathcal{T}_{\theta(a\,\cdot)}(\nu|\mu) \leqslant \mathcal{T}_{\theta_1(a\,\cdot)+\theta_2(a\,\cdot)}(\nu|\mu),$$

and we need to bound the right-hand side by the relative entropy to get a weak transport-entropy inequality. Obviously, the optimal weak cost on the right-hand side is lower-bounded by the sum of the weak costs, related to θ_1 and θ_2, but what we need is the reversed inequality. Fortunately, in dimension one, equality holds:

$$\overline{\mathcal{T}}_{\theta_1(a\,\cdot)+\theta_2(a\,\cdot)}(\nu|\mu) = \overline{\mathcal{T}}_{\theta_1(a\,\cdot)}(\nu|\mu) + \overline{\mathcal{T}}_{\theta_2(a\,\cdot)}(\nu|\mu). \tag{4}$$

We obtain this equality by showing that the two optimal weak transport costs $\overline{\mathcal{T}}_{\theta_1(a\,\cdot)}(\nu|\mu)$ and $\overline{\mathcal{T}}_{\theta_2(a\,\cdot)}(\nu|\mu)$ are achieved by the *same* coupling. This result is well known for classical transport cost $\mathcal{T}_\theta$ related to a convex cost function θ in dimension one. Namely, in the case where ν has no atom (for simplicity), the map

$$T_{\nu,\mu} = F_\mu^{-1} \circ F_\nu$$

is the only one non-decreasing and left-continuous function that pushes forward ν onto μ, that is to say

$$\int f\,d\mu = \int f \circ T_{\nu,\mu}\,d\nu.$$

Moreover, it follows from the works by Hoeffding, Fréchet and Dall'Aglio [10, 13, 20], that this map achieves the optimal transport of ν onto μ independently of the convex cost functions θ (see also [9]). In other words, it holds

$$\mathcal{T}_\theta(\mu,\nu) = \int \theta\left(|x - T_{\nu,\mu}(x)|\right)\,\nu(dx).$$

Actually, the expected equality (4) follows by combining this well known one dimensional statement with our following second main result.

Theorem 1.3. *Let $\mu,\nu \in \mathcal{P}_1(\mathbb{R})$; there exists a probability measure $\hat\gamma$ dominated by ν in the convex order, $\hat\gamma \preceq \nu$, such that* for all *convex cost function θ it holds*

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

We recall that $\nu_1 \preceq \nu_2$ means that $\int f\,d\nu_1 \leqslant \int f\,d\nu_2$ for all convex functions $f:\mathbb{R}\to\mathbb{R}$. As we shall see during the paper, this notion of convex ordering, characterized by Strassen [30] in terms of martingales, is really crucial for 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 [28] (see Theorem 2.8). From this geometrical interpretation, we obtain an intermediate outcome, Theorem 2.9, that can be interpreted as the discrete version of Theorem 1.3. Then, the proof of Theorem 1.3, given in Section 3, follows by discrete approximation arguments.

Let us come back to the main ideas of the proof of Theorem 1.2. The weak transport-entropy equality *i)* follows from condition *ii)* as follows; using equality (4), we get that

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

for a good choice of the constant a, by relating the condition *ii)*, either to a weak transport-entropy inequality with the cost function $\theta_2(a\,\cdot)$, either to a weak transport-entropy inequality with the cost function $\theta_1(a\,\cdot)$.

More precisely, adapting a preceding result by the first named author [15, Theorem 2.2], one shows in Theorem 6.1 that condition *ii)* characterizes the weak transport-entropy $\overline{\mathrm{T}}(\theta_2(a\,\cdot))$. Secondly, let us first observe that condition *ii)* implies that there exists $h>0$ such that

$$\sup_x \left(U_\mu(x+1) - U_\mu(x)\right) \leqslant h.$$

In their paper [5], Bobkov and Götze have shown that this weaker condition is equivalent to the Poincaré inequality (3) restricted to convex functions. In our last main result, we complete the picture by showing that this condition also characterizes measures satisfying a weak transport-entropy inequality with a cost function which is quadratic near zero and then linear, like $\theta_1(a\,\cdot\,)$.

Theorem 1.4. *Let μ be a probability measure on $\mathbb{R}$, then the following assertions are equivalent:*

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

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

(b) *There exists $C > 0$ such that for all convex function f on $\mathbb{R}$, μ satisfies*

$$\mathrm{Var}_\mu(f) \leqslant C\int_{\mathbb{R}} f'^2\,d\mu.$$

(c) *There exist $D, l_o > 0$ such that the probability μ satisfies the transport inequalities*

$$\overline{\mathcal{T}}_\alpha(\mu|\nu) \leqslant H(\nu|\mu), \qquad \forall \nu \in \mathcal{P}_1(\mathbb{R}),$$

and

$$\overline{\mathcal{T}}_\alpha(\nu|\mu) \leqslant H(\nu|\mu), \qquad \forall \nu \in \mathcal{P}_1(\mathbb{R}),$$

for the function α defined by

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

Moreover the constants are related as follows:

- $(a) \Rightarrow (c)$ *with $D = 2Kh^2$ and $l_o = c/h$,*
- $(a) \Rightarrow (b)$ *with $C = K'h^2$,*
- $(b) \Rightarrow (a)$ *with $h = K''\sqrt{h}$,*
- $(c) \Rightarrow (b)$ *with $C = D$,*

where c, K, K' and K'' are absolute constants.

We indicate that during the preparation of this work, we learned that this characterization of the convex Poincaré inequality in terms of transport-entropy inequality (actually in terms of their equivalent convex (τ) Property formulation) has also been obtained by Feldheim, Marsiglietti, Nayar and Wang in their recent paper [12].

The proof of Theorem 1.4 is given in Section 5. It uses results of independent interest like a new discrete logarithmic Sobolev inequality for the exponential measure τ (see Theorem 5.1). By transportation technics, this logarithmic-Sobolev inequality provides logarithmic-Sobolev inequalities restricted to the class of convex or concave functions for measures satisfying the condition (a) (see Corollary 5.2). Then the weak transport-entropy inequalities of Item (c) are obtained in their dual forms, involving infimum convolution operators (see Lemma 4.1). The method to derive transport-entropy inequalities from logarithmic-Sobolev inequalities is based on classical arguments involving the infimum convolution operator as a solution of the Hamilton-Jacobi equation. This approach is due to [4] and has been also generalized in [17, 18]. Finally, in the Appendix, we present a new discrete Hardy-type of inequality in Lemma 6.2. This is a useful tool to reach the discrete logarithmic Sobolev inequality for the exponential measure τ. Let us observe that this Lemma could also be used to recover differently the characterization by Bobkov and Götze, of the Poincaré inequality restricted to convex functions.

Acknowledgment: The authors would like to thank Greg Blekherman for discussions on the geometric aspects in Section 2.2, including his help with the proof of Theorem 2.9.

2. Convex ordering and a majorization lemma

2.1. **A reminder on convex ordering and Strassen's theorem.** We recall here some basic facts about convex ordering of probability measures and majorization of vectors. We refer to [22] and [19] for further results and bibliographic references.

Definition 2.1. Let ν_1, ν_2 in $\mathcal{P}_1(\mathbb{R})$; we say that ν_2 dominates ν_1 in the convex order if for all convex functions f on $\mathbb{R}$,

$$\int_{\mathbb{R}} f\, d\nu_1 \leqslant \int_{\mathbb{R}} f\, d\nu_2.$$

In this case we write $\nu_1 \preceq \nu_2$.

Let us remark that for probability measures belonging to $\mathcal{P}_1(\mathbb{R})$ the integral of convex functions always makes sense in $\mathbb{R} \cup \{+\infty\}$.

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

Proposition 2.2. *Let $\nu_1, \nu_2 \in \mathcal{P}_1(\mathbb{R})$; the following are equivalent:*

(1) $\nu_1 \preceq \nu_2$,

(2) $\int x\, \nu_1(dx) = \int x\, \nu_2(dx)$ *and for all Lipschitz and non-decreasing and non-negative convex function* $f : \mathbb{R} \to \mathbb{R}^+$,

$$\int f(x)\, \nu_1(dx) \leqslant \int f(x)\, \nu_2(dx).$$

(3) $\int x\, \nu_1(dx) = \int x\, \nu_2(dx)$ *and for all* $t \in \mathbb{R}$,

$$\int [x-t]_+\, \nu_1(dx) \leqslant \int [x-t]_+\, \nu_2(dx).$$

For the reader's convenience we sketch the proof of this classical result. We refer to [22] for more details.

Sketch of proof. Let us show that (1) is equivalent to (2). First of all, since the functions $x \mapsto x$ and $x \mapsto -x$ are convex, it is clear that $\nu_1 \preceq \nu_2$ implies that $\int x\, \nu_1(dx) = \int x\, \nu_2(dx)$ and so (1) implies (2). Conversely, since the graph of a convex function always lies above its tangent, subtracting an affine function if necessary, it is clear that one can restrict to non-negative convex functions. Moreover, if $f : \mathbb{R} \to \mathbb{R}$ is a convex function, then the function f_n defined by $f_n = f$ on $[-n, n]$, $f_n(x) = f_n(n) + f_n'(n)(x-n)$ if $x \geqslant n$ and $f_n(x) = f_n(-n) + f_n'(-n)(x+n)$ if $x \leqslant -n$ (where f_n' denotes for instance 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.

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 $\alpha_i \geqslant 0$ and $t_i \in \mathbb{R}$. This shows that (2) and (3) are equivalent. □

Let us also recall a classical result of Strassen [30] characterizing the convex ordering in terms of martingales.

Theorem 2.3. *Let $\nu_1, \nu_2 \in \mathcal{P}_1(\mathbb{R})$; the following are equivalent*

(1) $\nu_1 \preceq \nu_2$

(2) There exists a martingale (X, Y) such that X has law ν_1 and Y has law ν_2.

We refer to [18] for a proof of Theorem 2.3 involving Kantorovich duality for transport costs of the form $\overline{\mathcal{T}}$.

Convex ordering is closely related to the notion of majorization which is recalled in the following definition.

Definition 2.4. 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 each vector is equal.

Thus, assuming that the components of a and b are in increasing order, a is majorized by b, if

$$a_n + a_{n-1} + \cdots + a_{n-j+1} \leq b_n + b_{n-1} + \cdots + b_{n-j+1}, \quad 1 \leq j < n\,,$$

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

The following proposition recalls the link between majorization and convex ordering.

Proposition 2.5. *Let $a, b \in \mathbb{R}^n$ and denote by $\nu_1 = \frac{1}{n}\sum_{i=1}^n \delta_{a_i}$ and $\nu_2 = \frac{1}{n}\sum_{i=1}^n \delta_{b_i}$; the following are equivalent*

1. *(1) a is majorized by b,*
2. *(2) 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 $\sum_{i=1}^n f(a_i) \leq \sum_{i=1}^n f(b_i)$.*

In the sequel we will also denote $a \preceq b$ if a is majorized by b.

Proof. Let us show that (1) implies (2). We assume without loss of generality that the components of a and b are sorted in increasing order. Let us show that for all $t \in \mathbb{R}$, $\sum_{k=1}^n [a_k - t]_+ = \sum_{k=k_o}^n (a_k - t) \leqslant \sum_{k=1}^n [b_k - t]_+$. The case $t > \max a_k$ is clearly true. Take $t \leqslant \max a_k$ and let k_o be the smallest k such that $a_k \geqslant t$. Then it holds

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

where the first inequality comes from the majorization assumption. According to Proposition (2.2), this shows that $\nu_1 \preceq \nu_2$.

Now let us prove that (2) implies (1). First of all, the measures ν_1 and ν_2 have the same mean, and so $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i$. Let $k \in \{1, \ldots, n\}$; choosing $f(x) = [x - b_k]_+$, it holds

$$\sum_{i=k}^n a_i - b_k \leqslant \sum_{i=1}^n [a_i - b_k]_+ \leqslant \sum_{i=1}^n [b_i - b_k]_+ = \sum_{i=k}^n b_i - b_k,$$

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

Let us recall one simple classical consequence of Proposition 2.5 above in terms of discrete optimal transport on the line.

Lemma 2.6. *Let $x, y \in \mathbb{R}^n$ be two vectors such that $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$ and $y_1 \leqslant y_2 \leqslant \ldots \leqslant y_n$. Then for all permutation σ of $\{1, \ldots, n\}$ and all convex function $\theta : \mathbb{R} \to \mathbb{R}$, it holds*

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

Proof. Since, for all k, $\sum_{i=k}^n y_i \geqslant \sum_{i=k}^n y_{\sigma(i)}$, it holds for $\sum_{i=k}^n (x_i - y_i) \leqslant \sum_{i=k}^n (x_i - y_{\sigma(i)})$ (with equality for $k = 1$). Therefore, denoting $y_\sigma = (y_{\sigma(1)}, \ldots, y_{\sigma(n)})$, it holds $x - y \preceq x - y_\sigma$. Applying Proposition 2.5 completes the proof. □

Remark 2.7. In particular, let μ, ν are two discrete probability measures on $\mathbb{R}$ of the form

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

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.

The following characterization is due to Rado [28]. Here we deduce it from Strassen's Theorem. We will denote by $\mathcal{S}_n$ the set of all permutations of $\{1, 2, \ldots, n\}$.

Theorem 2.8. *Let $a, b \in \mathbb{R}^n$; the following are equivalent*

(1) The vector a is majorized by b,
(2) There exists a doubly stochastic matrix P such that $a = bP$ (treating a and b as row vectors),
(3) The vector a lies in the convex hull of the permutations of b, i.e. writing for all permutation $\sigma \in \mathcal{S}_n$, $b_\sigma = (b_{\sigma(1)}, \ldots, b_{\sigma(n)})$, it holds

$$a = \sum_{\sigma} \lambda_\sigma b_\sigma,$$

for some $\lambda_\sigma \geqslant 0$ with $\sum_\sigma \lambda_\sigma = 1$.

Proof. Let us prove that (1) implies (2). According to Proposition 2.5, $a \preceq b$ means 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. Let us write $\nu_1 = \frac{1}{n}\sum_{x\in\mathcal{X}} k_x \delta_x$ and $\nu_2 = \frac{1}{n}\sum_{y\in\mathcal{Y}} \ell_y \delta_y$, with $\mathcal{X} = \{a_1, \ldots, a_n\}$, $\mathcal{Y} = \{b_1, \ldots, b_n\}$ and k_x is the cardinal of the $\{i \in \{1, \ldots, n\} : a_i = x\}$ and ℓ_y the cardinal of the set $\{i \in \{1, \ldots, n\} : b_i = y\}$. According to Strassen's Theorem, 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 discrete,

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

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

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

with $K_{y,x} = n\frac{\mathbb{P}(X=x,Y=y)}{k_x \ell_y}$. Therefore, denoting $P_{j,i} = K_{b_j,a_i}$ for all i, j it holds $a = bP$ and P is doubly stochastic.

If $a = bP$ with a doubly stochastic P, then it is easily checked that $\sum_{i=1}^n f(a_i) \leqslant \sum_{i=1}^n f(b_i)$, for all convex function f on $\mathbb{R}$ and so (2) implies (1).

Finally, according to Birkhoff's theorem, 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. This shows that (2) and (3) are equivalent. ☐

2.2. **Geometric aspects.** In all what follows, we assume that the vector $b = (b_1, b_2, \ldots, b_n) \in \mathbb{R}^n$ has *distinct* components.

In the following, we will be working with the convex polytope denoted by $\mathrm{Perm}(b)$ and defined by

$$\mathrm{Perm}(b) := \mathrm{Conv}\{b_\sigma : \sigma \in \mathcal{S}_n\}$$

This polytope is sometimes called the *Permutahedron* generated by the vector b. According to Rado's Theorem 2.8, $\mathrm{Perm}(b)$ is exactly $\{a \in \mathbb{R}^n; a \preceq b\}$. This set is a subset of the 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 $\mathcal{E}_0 = \{x \in \mathbb{R}^n; \sum_{i=1}^n x_i = 0\}$.

In what follows, we are interested in the faces, facets containing a given face, and normal vectors to such facets.

Towards this, denote $[n]$ the set of integer 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). More generally, given $\mathcal{S} = (S_1, S_2, \dots, S_k)$, with $[n] = S_1 \cup S_2 \cup \dots \cup S_k$ being a partition, let $v_{\mathcal{S}}(b)$ denote the vector with the largest $|S_1|$ coordinates of b in the positions indexed by S_1, then the next largest $|S_2|$ coordinates in the positions indexed by S_2 and so on. Now recalling the following two fact from [3]:

Fact 1. Each *facet* of $\mathrm{Perm}(b)$ is the convex hull of the set $P_S(b)$ of points, for some $\emptyset \neq S \subset [n]$, where $P_S(b)$ contains the vector $v_S(b)$ along with all vectors obtained by permuting any subset of coordinates, 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 $[n] \setminus S$.)

Fact 2. More generally, let F be a face of $\mathrm{Perm}(b)$. Then F can be described as the convex hull of the set $P_{\mathcal{S}}(b)$ of points, for some $\mathcal{S} = (S_1, S_2, \dots, S_k)$ as described above, where $P_{\mathcal{S}}(b)$ contains 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. Furthermore, given a face F, with the description as above using an appropriate $\mathcal{S}$, the facets containing F can be obtained by coalescing the first several S_i's in $\mathcal{S}$ to obtain a partition into just two sets: that is, for each $1 \leq j \leq k-1$, facet $\mathcal{F}_j$ containing F can be described by taking $[n] = T_1 \cup T_2$, where $T_1 = S_1 \cup \dots \cup S_j$, and $T_2 = S_{j+1} \cup \dots \cup S_k$. (In particular, there are $k-1$ facets containing such a face.)

Using the above facts from (the geometry of) the theory of majorization, we now prove the following.

Theorem 2.9. *Let $a, b \in \mathbb{R}^n$ and assume that b has distinct coordinates and $a \notin \mathrm{Perm}(b)$. Then the following are equivalent:*

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

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

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

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

Moreover the vector $\hat{c}$ is sorted as a : $(a_i \leqslant a_j) \Rightarrow (\hat{c}_i \leqslant \hat{c}_j)$, 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

$$\langle a - \bar{c}, c - \bar{c} \rangle \leqslant 0, \qquad \forall c \in \mathrm{Perm}(b). \tag{5}$$

Proof. First of all, we can assume that $\sum_{i=1}^n a_i = \sum_{i=1}^n b_i$. Indeed if this is not the case, 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 (5)) that the orthogonal projection of a and $\tilde{a}$ on $\mathrm{Perm}(b)$ are equal (to some point denoted by $\hat{c}$), and that $a - \hat{c} \preceq a - c$ if and only if $\tilde{a} - \hat{c} \preceq \tilde{a} - c$.

$(i) \implies (ii)$. Let $\bar{c}$ (not necessarily equal to the $\hat{c}$ in (i)) satisfy (ii).

Then by (i),

$$a - \hat{c} \preceq a - \bar{c},$$

which, by Proposition 2.5 implies that $\sum_{i=1}^n (a-\hat{c})_i^2 \le \sum_{i=1}^n (a-\bar{c})_i^2$; by the choice of $\bar{c}$, this is possible only if $\hat{c} = \bar{c}$.

$(ii) \implies (i)$. Let $\hat{c}$ be defined as in (ii) of the proposition. Since $\mathrm{Perm}(b)$ is invariant by permutation, it easily follows from Lemma 2.6 that the coordinates of $\hat{c}$ are in the same order as the coordinates of a.

Now let us prove that $a - \hat{c} \preceq a - c$ for all $c \in \mathrm{Perm}(b)$. 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.

(a) A simple case. 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 obtained by permuting two coordinates of x (corresponding to indices both in S or 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 value of α on these sets, which verify also $k\alpha_S + (n-k)\alpha_{S^c} = 0$ since $\alpha \in \mathcal{E}_0$

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

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

This amounts to showing that

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

Indeed, the choice of $\hat{c}$ implies that we have $\langle \alpha, \hat{c}\rangle \ge \langle \alpha, c'\rangle$, for every $c' \in \mathrm{Perm}(b)$, hence $\langle \alpha, c' - \hat{c}\rangle \le 0$; the second condition follows, since $\hat{c}, c' \in \mathrm{Perm}(b)$ implies $\sum_i c_i = \sum_i c'_i - \sum_i \hat{c}_i = 0$.

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

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

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 $c_S \leqslant 0$ and $c_{S^c} \geqslant 0$. If f is a convex function on $\mathbb{R}$, then, according to Jensen inequality

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

where the last inequality comes from the fact that

$$f'(\alpha_S) c_S + f'(\alpha_S^c) c_{S^c} = c_S (f'(\alpha_S) - f'(\alpha_{S^c})) \leqslant 0.$$

According to Proposition 2.5, we conclude that $\alpha \preceq \alpha - c$.

(b) The general case. Suppose that $\hat{c}$ lies in a face F of the polytope. This face is related to a partition $\mathcal{S} = (S_1, \ldots, S_k)$ of $[n]$. 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 $i \in \{1, \ldots, n-1\}$, let us denote by F_i the facet containing F associated to the partition $\mathcal{T}_i = \{S_1 \cup \ldots \cup S_i\,;\, S_{i+1} \cup \ldots \cup S_k\}$, $1 \leqslant i \leqslant k-1$. Consider the vectors $p_1, p_2, \ldots, p_{k-1} \in \mathcal{E}_0$ 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 $\langle c, p_i \rangle \leqslant \langle v_{\mathcal{T}_i}, p_i \rangle$, 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 $i \in S_j$, $\alpha_i = \left(\sum_{p=j}^{k-1} \lambda_p\right) - \sigma := A_j$.

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:

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

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

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

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

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

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

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

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

which proves that $a \preceq a - d$. □

3. Properties of the optimal coupling for weak transport costs

It is well known that if μ and ν are two probability measures on $\mathbb{R}$ and U is a random variable uniformly distributed on $[0,1]$, the coupling

$$\pi^* = \mathrm{Law}(F_\mu^{-1}(U), F_\nu^{-1}(U)),$$

is optimal for all transport costs $\mathcal{T}_\theta(\mu,\nu)$ associated to a convex cost function θ. (This fact generalizes the result mentioned in Remark 2.7). As a consequence, on $\mathbb{R}$, the transport cost is additive : for all convex cost functions α and β,

$$\mathcal{T}_{\alpha+\beta}(\mu,\nu) = \mathcal{T}_\alpha(\mu,\nu) + \mathcal{T}_\beta(\mu,\nu).$$

In this section, we will establish a similar result (Corollary 3.3) for the weak transport costs $\overline{\mathcal{T}}$.

Toward this, we begin with the following proposition which illustrates a relationship between $\mathcal{T}$ and $\overline{\mathcal{T}}$. In the sequel, we denote by $\mathrm{Im}(\mu)$ the set of probability measures on $\mathbb{R}$ which are images of μ under some map $S : \mathbb{R} \to \mathbb{R}$, *i.e.*

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

and by $\mathrm{Im}^\uparrow(\mu)$ the set of probability measures which are images of μ under a non-decreasing map S.

Proposition 3.1. *For all probability measures μ, ν on $\mathbb{R}$, it holds*

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

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 1.3 below, we will use Proposition 3.1 with μ being a uniform distribution on n distinct points. In this case, it is clear that $\mathrm{Im}^\uparrow(\mu) = \mathrm{Im}(\mu)$.

Proof. Firstly, we prove that $\overline{\mathcal{T}}_\theta(\nu|\mu) \geqslant \inf_{\gamma \preceq \nu, \gamma \in \mathrm{Im}(\mu)} \mathcal{T}_\theta(\gamma,\mu)$. Namely, let $\pi(dxdy) = p(x,dy)\mu(dx)$ be some coupling between μ and ν and let us denote by $S(x) = \int y\, p(x,dy)$, $x \in \mathbb{R}$. Clearly $S_\#\mu \in \mathrm{Im}(\mu)$. Moreover $S_\#\mu \preceq \nu$. Indeed if $f : \mathbb{R} \to \mathbb{R}$ is some convex function, then by Jensen inequality, it holds

$$\int f(x)\, S_\#\mu(dx) = \int f\left(\int y\, p(x,dy)\right) \mu(dx) \leqslant \iint f(y)\, p(x,dy)\mu(dx) = \int f(y)\, \nu(dy).$$

Then it holds

$$\int \theta\left(x - \int y\, p(x,dy)\right) \mu(dx) = \int \theta(x - S(x))\, \mu(dx) \geqslant \mathcal{T}_\theta(S_\#\mu, \mu) \geqslant \inf_{\gamma \preceq \nu, \gamma \in \mathrm{Im}(\mu)} \mathcal{T}_\theta(\gamma,\mu).$$

Therefore taking the infimum over p yields to the desired inequality.

Now we turn to the proof of the inequality $\overline{\mathcal{T}}_\theta(\nu|\mu) \leqslant \inf_{\gamma\preceq\nu,\gamma\in\mathrm{Im}^\uparrow(\mu)} \mathcal{T}_\theta(\gamma,\mu)$. 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 $\pi_1(dxdy) = p_1(x,dy)\gamma(dx)$ and $x = \int_\mathbb{R} y p_1(x,dy)$, γ almost everywhere. For all $x \in \mathbb{R}$, define probability measure $p(x,dy) := p_1(S(x),dy)$. Then for all bounded continuous function f, it holds

$$\iint f(y)p(x,dy)\,\mu(dx) = \iint f(y)p_1(S(x),dy)\,\mu(dx)$$
$$= \iint f(y)p_1(x,dy)\,\gamma(dx) = \int f(y)\,\nu(dy).$$

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

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

Since S is non-decreasing, it realizes the optimal transport between μ and ν for the cost $\mathcal{T}_\theta$ and so it follows that

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

□

We now turn to the proof of Theorem 1.3.

Let us emphasize that the probability $\hat\gamma$ depends on μ and ν but not on the cost function θ. This will be crucial in the sequel.

Proof of Theorem 1.3.

Step 1. We first treat the case where

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

with $a_1 < a_2 < \ldots < a_n$ and $b_1 < b_2 < \ldots < b_n$. We denote by $a = (a_1,\ldots,a_n)$ and $b = (b_1,\ldots,b_n)$. According to Theorem 2.9, 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 $\hat c_i \leqslant \hat c_{i+1}$. Let us denote by $\hat\gamma = \frac{1}{n}\sum_{i=1}^n \delta_{\hat c_i}$. The probability ν dominates $\hat\gamma$ for the stochastic order and if $\theta : \mathbb{R}^+ \to \mathbb{R}^+$ is a convex cost function, it holds

$$\frac{1}{n}\sum_{i=1}^n \theta(|a_i - \hat c_i|) \leqslant \inf_{c\in\mathrm{Perm}(b)} \frac{1}{n}\sum_{i=1}^n \theta(|a_i - c_i|). \tag{6}$$

Since the coordinates of $\hat c$ are non-decreasing,

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

On the other hand, according to Proposition 3.1,

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

(here we use that for such a distribution μ, it holds $\mathrm{Im}(\mu) = \mathrm{Im}^\uparrow(\mu)$). 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 $c_i \leqslant c_{i+1}$ and $c = (c_1,\ldots,c_n) \in \mathrm{Perm}(b)$, and for such a c, it holds $\mathcal{T}_\theta(\gamma,\mu) = \frac{1}{n}\sum_{i=1}^n \theta(|a_i - c_i|)$. So, using (6) we conclude that

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

Step 2. In this second step, we treat the general case using a natural approximation argument.

Let μ and ν be two elements of $\mathcal{P}_1(\mathbb{R})$. By assumption, $\int |x|\,\mu(dx) < \infty$ and $\int |x|\,\nu(dx) < \infty$, so, according to the de la Vallée-Poussin theorem, 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 $\int \beta(|x|)\,\nu(dx) < \infty$.

Now let us construct discrete approximations of μ and ν. According to Varadarajan's theorem, 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 ∞. Let us take $(x_i)_{i\geqslant 1}$ a positive realization of these events and set $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i^{(n)}}$, where $x_1^{(n)} \leqslant x_2^{(n)} \leqslant \ldots \leqslant x_n^{(n)}$ denotes the increasing re-ordering of the vector $(x_1, x_2, \ldots, x_n)$. Then the sequence μ_n converges weakly to μ and $\int |x|\,\mu_n(dx) \to \int |x|\,\mu(dx)$. According to Theorem 6.9 of [32], this is equivalent to the convergence for 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 $n \geqslant 1$, 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 function θ. Let us show that one can extract from $\hat{\gamma}_n$ a sub-sequence 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 $M = \sup_{n\geqslant 1} \int \beta(|x|)\,\nu_n(dx)$ is finite. Since $\hat{\gamma}_n \preceq \nu_n$ and since the function $x \mapsto \beta(|x|)$ is convex, it thus holds $\int \beta(|x|)\,\gamma_n(dx) \leqslant \int \beta(|x|)\,\nu_n(dx) \leqslant M$. In particular, setting $c(R) = \inf_{t\geqslant R} \beta(t)/t$, $R > 0$, Markov's inequality easily implies that

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

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,

$$\sup_{n\geqslant 1} \tilde{\gamma}_n([-R,R]^c) \leqslant \frac{2M}{c(R)}, \qquad \forall R \geqslant 1$$

and so the sequence $\tilde{\gamma}_n$ is tight. Therefore, according to 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

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

where $\hat{\gamma}(dx) = \frac{Z}{1+|x|}\,\tilde{\gamma}(dx) \in \mathcal{P}_1(\mathbb{R})$. According to Theorem 6.9 of [32], this implies $\hat{\gamma}_n \to \hat{\gamma}$ as $n \to \infty$ for the W_1 distance.

Now let us check that $\hat{\gamma}$ is such that $\overline{\mathcal{T}_\theta}(\nu|\mu) = \mathcal{T}_\theta(\hat{\gamma}, \mu)$ for all convex cost function $\theta : \mathbb{R}^+ \to \mathbb{R}^+$. First we assume that θ is Lipschitz, and we denote by L_θ the Lipschitz constant of θ. According to Theorem 2.11 of [18], 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 $\bar{\varphi} \leqslant \varphi$, $\bar{\varphi}$ 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 function and the fact that $Q_\theta\varphi$ is L_θ-Lipschitz (being an infimum of such functions), we easily get the following inequality

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

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

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

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

Now we want 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 $n \geqslant 1$, 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 $x \geqslant n$, 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 $\int Q_\theta\varphi\, d\gamma = \sup_{n\geqslant 1}\int Q_{\theta_n}\varphi\, d\gamma$. We deduce from this that $\overline{\mathcal{T}}_\theta(\nu|\mu) = \sup_{n\geqslant 1}\overline{\mathcal{T}}_{\theta_n}(\nu|\mu)$ and $\mathcal{T}_\theta(\hat{\gamma}|\mu) = \sup_{n\geqslant 1}\mathcal{T}_{\theta_n}(\hat{\gamma}, \mu)$. Since $\overline{\mathcal{T}}_{\theta_n}(\nu|\mu) = \mathcal{T}_{\theta_n}(\hat{\gamma}, \mu)$ for all $n \geqslant 1$, this ends the proof. □

Corollary 3.3. *Let α and β be two convex cost functions, then for all probability measures $\mu, \nu \in \mathcal{P}_1(\mathbb{R})$, it holds*

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

Proof of Corollary 3.3. According to Corollary 1.3, 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 functions $\theta \in \{\alpha, \beta, \alpha + \beta\}$. The result then follows from the additivity of $\mathcal{T}_\theta$ in dimension one :

$$\mathcal{T}_{\alpha+\beta}(\hat{\gamma}, \mu) = \mathcal{T}_\alpha(\hat{\gamma}, \mu) + \mathcal{T}_\beta(\hat{\gamma}, \mu).$$

□

4. Dual formulation for weak transport-entropy inequalities.

Let us recall the dual formulations in terms of infimum convolution inequalities of the transport-entropy inequality $\mathrm{T}(\theta)$ defined at (1) and the weak transport-entropy inequalities presented in Definition 1.1. The following result is stated in dimension one only, but its conclusion is very general (see for instance [18]).

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 function $g : \mathbb{R} \to \mathbb{R}$ bounded from below, consider*

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

(1) *The probability measure* μ *satisfies* $\mathrm{T}(\theta)$ *if and only if*

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

for all continuous function $g : \mathbb{R} \to \mathbb{R}$ *bounded from below.*

(2) *The probability measure* μ *satisfies* $\overline{\mathrm{T}}^{+}(\theta)$ *if and only if*

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

for all convex function $g : \mathbb{R} \to \mathbb{R}$ *bounded from below.*

(3) *The probability measure* μ *satisfies* $\overline{\mathrm{T}}^{-}(\theta)$ *if and only if*

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

for all convex function $g : \mathbb{R} \to \mathbb{R}$ *bounded from below.*

(4) *If* μ *satisfies* $\overline{\mathrm{T}}(\theta)$*, then it satisfies*

$$\int \exp(Q_t g)\, d\mu \int \exp(-g)\, d\mu \leqslant 1, \tag{7}$$

with $t = 2$.

Conversely, if μ *satisfies* (7) *for some* $t > 0$*, then it satisfies* $\overline{\mathrm{T}}(t\theta(\,\cdot\,/t))$.

Proof. The first item is due to Bobkov and Götze [6]. The proof is an easy combination of well known duality formulas for the relative entropy and for the transport cost $\mathcal{T}_\theta$ given by Kantorovich duality theorem (see *e.g* [32]). Items (2) and (3) generalize the first point to the framework of weak transport-entropy inequalities. We refer to Proposition 4.5 of [18] for a more general statement and a proof (based on an extension of Kantorovich duality to a more general optimal transport costs also obtained in [18]). Let us sketch the proof of Item (4) (which already appeared in a slightly different form in [16] Propositions 8.2 and 8.3). By definition if μ satisfies $\overline{\mathrm{T}}(\theta)$ then it satisfies $\overline{\mathrm{T}}^{\pm}(\theta)$. Therefore, it satisfies the exponential inequalities given in Items (2) and (3). Note that if g is convex and bounded from below then $Q_1 g$ is also convex and bounded from below. Therefore it holds

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

and

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

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* Theorem 22.46 of [32]) gives (7). The converse implication simply follows from Jensen inequality. □

5. A transport form of the convex Poincaré inequality

This section is devoted to the proof of Theorem 1.4. In this statement, from [5], we have the equivalence between (a) and (b), and it is easy to prove $(c) \Rightarrow (b)$. To complete the proof, we thus have to show that if μ satisfies (a) and (b) it also satisfies (c). Our strategy is the following. We will first prove a modified logarithmic Sobolev inequality for the exponential probability measure τ. Considering a probability measure μ satisfying Items (a) (and (b)) above, we will then deduce (using a transport argument involving the map U_μ) a modified logarithmic Sobolev inequality for μ restricted to convex or concave Lipschitz functions. This last property will finally imply (following the well known Hamilton-Jacobi interpolation technique of [4]) the desired transport inequality.

First, let us state the modified logarithmic Sobolev inequality for the exponential probability measure τ.

Theorem 5.1. *There exist $K > 0$ and $1 > c > 0$ such that for all non-decreasing function f with $f(x) - f(x-1) \leqslant c$ for all x, it holds*

$$\mathrm{Ent}_\tau(e^f) \leqslant K \int_{\mathbb{R}} (f(x) - f(x-1))^2 \, e^f \, d\tau. \tag{8}$$

and

$$\mathrm{Ent}_\tau(e^f) \leqslant K \int_{\mathbb{R}} (f(x+1) - f(x))^2 \, e^f \, d\tau. \tag{9}$$

This result is a variant of an analogous result by Bobkov and Ledoux [7]. The main difference between the two results is that in the version given above, the entropy of e^f is controlled by some integral involving the *discrete* gradient of f instead of the usual first derivative of f. The proof of Theorem 5.1 adapts different arguments appearing in [7]. The proof is postponed to the Appendix.

Now let us deduce from Theorem 5.1 modified logarithmic Sobolev inequalities for probability measures satisfying Item (a) of Theorem 1.4. In what follows, if g is a convex or a concave function on $\mathbb{R}$, we will use the notation $|\nabla g|$ to denote the function defined by

$$|\nabla g|(x) = \min\{|\theta g'_-(x) + (1-\theta) g'_+(x)|; \theta \in [0,1]\}, \tag{10}$$

where g'_- and g'_- denote the left and right derivatives of the function g (which are well defined everywhere). So in the convex case, it holds

$$|\nabla g|(x) = \min(|t|; t \in [g'_-(x), g'_+(x)]) = \begin{cases} |g'_+(x)| & \text{if } g'_+(x) \leqslant 0 \\ 0 & \text{if } g'_-(x) \leqslant 0 \leqslant g'_+(x) \\ g'_-(x) & \text{if } g'_-(x) \geqslant 0 \end{cases}$$

and in the concave case

$$|\nabla g|(x) = \min(|t|; t \in [g'_+(x), g'_-(x)]) = \begin{cases} |g'_-(x)| & \text{if } g'_-(x) \leqslant 0 \\ 0 & \text{if } g'_+(x) \leqslant 0 \leqslant g'_-(x) \\ g'_+(x) & \text{if } g'_+(x) \geqslant 0 \end{cases}$$

Corollary 5.2. *Assume that μ satisfies Item (a) of Theorem 1.4 with a constant $h > 0$, then for all convex or concave and l-Lipschitz function g with $l \leqslant c/h$, it holds*

$$\mathrm{Ent}_\mu(e^g) \leqslant K h^2 \int_{\mathbb{R}} |\nabla g(x)|^2 e^{g(x)} \mu(dx), \tag{11}$$

where the constants K and c are those appearing in Theorem 5.1.

Proof. Case of convex functions. We will first prove (11) for a convex *non-decreasing* and l-Lipschitz function g with $l \leqslant c/h$. Set $f = g \circ U_\mu$, then

$$f(x) - f(x-1) \leqslant g(U_\mu(x)) - g(U_\mu(x) - h) \leqslant lh \leqslant c,$$

for all $x \in \mathbb{R}$. Since μ is the image of τ under the map U_μ, applying (8) to f leads to

$$\mathrm{Ent}_\mu(e^g) \leqslant K \int_{\mathbb{R}} (g(x) - g(x-h))^2 \, e^{g(x)} \mu(dx) \leqslant K h^2 \int_{\mathbb{R}} g'_-(x)^2 e^{g(x)} \mu(dx),$$

where the second inequality is due to the fact that g is convex and non-decreasing and therefore satisfies $0 \leqslant g(x) - g(x-h) \leqslant g'_-(x) h$.

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

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

Applying (8) to this function f and using that τ is *symmetric* and that U_μ transforms τ into μ yields to

$$\mathrm{Ent}_\mu(e^g) \leqslant K \int_{\mathbb{R}} (g(x) - g(x+h))^2 \, e^{g(x)} \mu(dx) \leqslant K h^2 \int_{\mathbb{R}} g'_+(x)^2 e^{g(x)} \mu(dx),$$

where we used that $0 \leqslant g(x) - g(x+h) \leqslant g'_+(x)(-h)$.

Finally, let us consider an arbitrary convex and l-Lipschitz function g with $l \leqslant c/h$. We can assume 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 $g(a) = 0$. 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, monotonic and l-Lipschitz. Therefore, according to what precedes, it holds

$$\mathrm{Ent}_\mu(e^{g_1}) \leqslant Kh^2 \int_{(-\infty,a]} {g_1}'_+(x)^2 e^{g(x)}\,\mu(dx) = Kh^2 \int_{(-\infty,a]} |\nabla g(x)|^2 e^{g(x)}\,\mu(dx)$$

and

$$\mathrm{Ent}_\mu(e^{g_2}) \leqslant Kh^2 \int_{[a,+\infty)} {g_2}'_-(x)^2 e^{g(x)}\,\mu(dx) = Kh^2 \int_{(a,+\infty)} |\nabla g(x)|^2 e^{g(x)}\,\mu(dx),$$

where we used that ${g_1}'_+(a) = {g_2}'_-(a) = |\nabla g(a)| = 0$. So all what remains to prove is the following subbadditivity property of the entropy functional:

$$\mathrm{Ent}_\mu(e^{g_1+g_2}) \leqslant \mathrm{Ent}_\mu(e^{g_1}) + \mathrm{Ent}_\mu(e^{g_2}).$$

Since $\int ge^g\,d\mu = \int g_1 e^{g_1}\,d\mu + \int g_2 e^{g_2}\,d\mu$ it is enough to prove that

$$\int e^g\,d\mu \log\left(\int e^g\,d\mu\right) \geqslant \int e^{g_1}\,d\mu \log\left(\int e^{g_1}\,d\mu\right) + \int e^{g_2}\,d\mu \log\left(\int e^{g_2}\,d\mu\right).$$

Set $t = \mu(-\infty, a]$ and $A = \frac{\int_{-\infty}^a e^g\,d\mu}{t} \geqslant 1$ and $x = \frac{\int_a^{+\infty} e^g\,d\mu}{1-t} \geqslant 1$, then the inequality above amounts to prove that for all $x \geqslant 1$

$$\varphi(x) := (tA + (1-t)x)\log(tA + (1-t)x) \geqslant \\ (tA + (1-t))\log(tA + (1-t)) + (t + (1-t)x)\log(t + (1-t)x) := \psi(x)$$

Since $\varphi(1) = \psi(1)$, it is enough to compare the first derivatives of φ and ψ. But

$$\varphi'(x) = (1-t)[1 + \log(tA + (1-t)x)]$$

and

$$\psi'(x) = (1-t)[1 + \log(t + (1-t)x)].$$

and since $A \geqslant 1$, it holds $\varphi'(x) \geqslant \psi'(x)$ which completes the proof.

Case of concave functions. Now assume that g is a differentiable, concave, non-decreasing and l-Lipschitz function with $l \leqslant c/h$. Set $f = g \circ U_\mu$, then $f(x+1) - f(x) \leqslant g(U_\mu(x) + h) - g(U_\mu(x)) \leqslant lh \leqslant c$. Using this time (9) yields to

$$\mathrm{Ent}_\mu(e^g) \leqslant K \int_{\mathbb{R}} (g(x+h) - g(x))^2\, e^{g(x)} \mu(dx) \leqslant Kh^2 \int_{\mathbb{R}} g'_+(x)^2 e^{g(x)} \mu(dx),$$

where the second inequality uses the fact that g is non-decreasing and concave. The extension of this inequality to all concave l-Lipschitz functions follows exactly the same line as what precedes. (Note that if $x, A \leqslant 1$, then $\varphi(1) = \psi(1)$ and $\varphi'(x) \leqslant \psi'(x)$ which implies that $\varphi(x) = \varphi(1) - \int_x^1 \varphi'(u)\,du \geqslant \psi(x)$.) □

Proof of Theorem 1.4. Let us show that (a) implies (c). The proof of this implication closely follows the proof of Corollary 5.1 in [4], so we will only sketch the main arguments and refer to [4, 17, 18] for details. According to Corollary 5.2, under (a) the inequality (11) holds for all convex or concave differentiable function g which is l_o-Lipschitz with $l \leqslant c/h := l_o$. Consider the function α defined by

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

whose convex conjugate functions $\alpha^*(v) = \sup_u\{uv - \alpha(u)\}$ satisfies

$$\alpha^*(v) = \begin{cases} Kh^2v^2 & \text{if } |v| \leqslant l_o \\ +\infty & \text{if } |v| > l_o \end{cases}.$$

With these notations (11) reads

$$\mathrm{Ent}_\mu(e^g) \leqslant \int \alpha^*(|\nabla g|)e^g \, d\mu, \tag{12}$$

for all convex or concave function $g : \mathbb{R} \to \mathbb{R}$.

Introduce the operators Q_t $t \in (0,1]$ defined by

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

which makes sense for instance for any l_o-Lipschitz function f or for any function f bounded from below. The operators Q_t have the following important properties. Let f be an l_o-Lipschitz function, then the following holds:

- If f is convex, then $Q_t f$ is convex: the property is easy to check and is actually a general fact about infimum convolution of two convex functions,
- If f is concave, then $Q_t f$ is concave: namely

$$Q_t f(x) = \inf_u \left\{ f(x-u) + t\alpha\left(\frac{u}{t}\right)\right\} \tag{13}$$

  and so $Q_t f$ is an infimum of concave functions and is therefore also concave.
- The function $Q_t f$ is l-Lipschitz: namely according to (13), it is an infimum of l-Lipschitz functions. Note that if f is assumed to be bounded from below, then, as observed above $Q_t f$ is well defined, and since the functions $x \mapsto t\alpha\left(\frac{x-y}{t}\right)$ are l_o-Lipschitz, the function $Q_t f$ is also l_o-Lipschitz.
- The function $u(t,x) = Q_t f(x)$ satisfies the following Hamilton-Jacobi equation:

$$\frac{d}{dt_+}u(t,x) + \alpha^*(|\nabla^- u|)(t,x) = 0, \tag{14}$$

  for all $t \in (0,1]$ and all $x \in \mathbb{R}$, where d/dt_+ is the right time derivative and

$$|\nabla^- u(t,x)| = \limsup_{y\to x} \frac{[u(t,y) - u(t,x)]_-}{|y-x|}.$$

  For this last point we refer to [17] or [1].

Let f be a convex function bounded from below or a concave and l-Lipschitz function with $l \leqslant l_o$; then following [4], define

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

The function F is right differentiable at every point $t > 0$ and it holds (see *e.g.* [17] for details)

$$\begin{aligned} F'(t) &= \frac{1}{t^2}\frac{1}{\int_{\mathbb{R}} e^{tQ_t f}\, d\mu}\left(\mathrm{Ent}_\mu\left(e^{tQ_t f}\right) + t^2\int_{\mathbb{R}} \frac{\partial}{\partial t}Q_t f e^{tQ_t f}\, d\mu\right) \\ &= \frac{1}{t^2}\frac{1}{\int_{\mathbb{R}} e^{tQ_t f}\, d\mu}\left(\mathrm{Ent}_\mu\left(e^{tQ_t f}\right) - Kh^2t^2\int_{\mathbb{R}} |\nabla^- Q_t f|^2 e^{tQ_t f}\, d\mu\right) \\ &\leqslant \frac{Kh^2}{\int_{\mathbb{R}} e^{tQ_t f}\, d\mu}\left(\int |\nabla Q_t f|^2 e^{tQ_t f}\, d\mu - \int_{\mathbb{R}} |\nabla^- Q_t f|^2 e^{tQ_t f}\, d\mu\right) \\ &\leqslant 0, \end{aligned}$$

where the second equality follows from (14), the first inequality from (12) applied to the function $g = tQ_t f$ (which is convex or concave and tl_o-Lipschitz) and the second inequality from the fact that for a convex or concave function g, $|\nabla g| \leqslant |\nabla^- g|$, where $|\nabla g|$ was defined at (10).

The function F is thus non-increasing and so it satisfies $F(1) \leqslant \lim_{t\to 0} F(t) = \int f\,d\mu$. In other words,

$$\int e^{Q_1 f}\,d\mu \leqslant e^{\int f\,d\mu}, \tag{15}$$

for any convex function bounded from below or for any concave and l-Lipschitz function f with $l \leqslant l_o$. In particular, restricting (15) to convex functions bounded from below and according to Point 3 of Lemma 4.1, one concludes that μ satisfies the transport inequality $\overline{\mathrm{T}}^-(\alpha)$:

$$\overline{\mathcal{T}}_\alpha(\mu|\nu) \leqslant H(\nu|\mu), \qquad \forall \nu \in \mathcal{P}_1(\mathbb{R}).$$

Applying the inequality to $f = -Q_1 g$ with g a convex function bounded from below (this function f is concave and l_o-Lipschitz) yields to

$$e^{\int Q_1 g\,d\mu} \int e^{Q_1(-Q_1 g)}\,d\mu \leqslant 1.$$

It is easy to check that $Q_1(-Q_1 f) \geqslant -f$ and so it holds

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

for all convex function g bounded from below. According to Point 2 of Lemma 4.1, this implies that μ satisfies the transport inequality $\overline{\mathrm{T}}^+(\alpha)$:

$$\overline{\mathcal{T}}_\alpha(\nu|\mu) \leqslant H(\nu|\mu), \qquad \forall \nu \in \mathcal{P}_1(\mathbb{R}).$$

This completes the proof. □

6. Proof of Theorem 1.2

First, we treat the particular case where the cost function vanishes on a neighborhood of 0.

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 propositions are equivalent:*

(1) There is $a > 0$ such that μ satisfies the transport-entropy inequality $\mathrm{T}(\beta(a\,\cdot))$.
(2) There is $a' > 0$ such that μ satisfies the weak transport-entropy inequality $\overline{\mathrm{T}}(\beta(a'\,\cdot))$.
(3) There are $b > 0$ and $K > 0$ such that $\max(K^+(b), K^-(b)) \leqslant K$, where

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

and

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

where m is a median of μ. (Here we use the convention $0/0 = 0$.)
(4) There is $d > 0$ such thatw

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

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

In particular, (2) implies (4) with $d = a'\frac{t_o}{8\beta^{-1}(\log 3)}$ and (4) implies (2) with $a' = d\frac{t_o}{9\beta^{-1}(2)}$.

Proof of Theorem 6.1. The equivalence between assertions (1), (3) and (4) was first proved in [15] (see Theorem 2.2). Let us complete the proof of Theorem 6.1 by showing that (1) ⇒ (2) ⇒ (3).

First of all, it follows easily from Jensen inequality that

$$\mathcal{T}_{\beta(a\,\cdot)}(\mu,\nu) \geqslant \max\left(\mathcal{T}_{\beta(a\,\cdot)}(\nu|\mu); \mathcal{T}_{\beta(a\,\cdot)}(\mu|\nu)\right).$$

Therefore (1) implies (2) with $a' = a$.

Now let us show that (2) implies (3). Suppose that μ satisfies $\overline{\mathrm{T}}(\beta(a\,\cdot))$ for some $a > 0$. According to Item (4) of Lemma 4.1, for all convex function $g : \mathbb{R} \to \mathbb{R}$ bounded from below, it holds

$$\int \exp(Qf)\, d\mu \int e^{-f}\, d\mu \leqslant 1,$$

where

$$Qf(x) = \inf_{y\in\mathbb{R}}\{f(y) + 2\beta(a|y-x|/2)\}.$$

Consider the convex function f_x which equals to 0 on $(-\infty, x]$ and ∞ otherwise, then $Qf(y) = 0$ on $(-\infty, x]$ and $Qf(y) = 2\beta(a(y-x)/2)$ on (x,∞). Applying the inequality above to f_x thus yields

$$\left(\mu(-\infty, x] + \int_{(x,\infty)} e^{2\beta(a(y-x)/2)}\mu(dy)\right)\mu(-\infty, x] \leqslant 1.$$

Considering $x \geqslant m$ yields that $K^+(a/2) \leqslant 3$. One proves similarly that $K^-(a/2) \leqslant 3$. This shows that (2) implies (3) with $b = a/2$ and $K = 3$. □

We are now in position to prove Theorem 1.2.

Proof of Theorem 1.2. 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 $\theta_2(t) = [\theta(t) - t^2]_+$. Note that θ_1 and θ_2 are both convex and that θ_2 vanishes on $[0, t_o]$ and that $\max(\theta_1, \theta_2) \leqslant \theta \leqslant \theta_1 + \theta_2$.

First assume that μ satisfies the weak transport-entropy inequality $\overline{\mathrm{T}}(\theta(a\,\cdot))$ for some $a > 0$. Then, since $\theta \geqslant \theta_2$ 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:

$$\sup_{x\in\mathbb{R}} U_\mu(x+u) - U_\mu(x) \leqslant \frac{1}{b}\theta_2^{-1}(u), \qquad \forall u > 0, \tag{16}$$

with $b = a\kappa$, where $\kappa = t_o/(8\theta_2^{-1}(\log 3)$. Since $\theta_2^{-1}(u) = \theta^{-1}(u + t_o^2)$ this proves the claim.

Now let us assume that μ satisfies (16) for some $b > 0$ and let us show that it satisfies $\overline{\mathrm{T}}(\theta(a\,\cdot))$ for some $a > 0$. First of all, taking $u = 1$ and using Theorem 1.4, one concludes that μ satisfies $\overline{\mathrm{T}}(\bar\alpha)$ with α defined by $\alpha(u) = \bar\alpha(u/\sqrt{2D})$, with $D = 2K\left(\theta^{-1}(1+t_o^2)\right)^2 \frac{1}{b^2}$ and

$$\bar\alpha(v) = \begin{cases} v^2 & \text{if } |v| \leqslant c\sqrt{K} \\ c\sqrt{2K}v - c^2K & \text{if } |v| > c\sqrt{K} \end{cases}$$

It is not difficult to check that $\bar\alpha$ compares to θ_1: for all $v \in \mathbb{R}$,

$$\bar\alpha(v) \geqslant \theta_1\left(\max(c\sqrt{K}/t_o; 1)v\right).$$

Therefore one concludes that μ satisfies $\overline{\mathrm{T}}(\theta_1(a_1\,\cdot))$, with

$$a_1 = \frac{\max((c\sqrt{K})/t_o; 1)}{2\sqrt{K}\theta^{-1}(1+t_o^2)}b.$$

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 $\overline{\mathrm{T}}(\theta_2(a\,\cdot))$, so since $\theta(at) \leqslant \theta_1(at) + \theta_2(at)$ and according to Corollary 3.3, it holds

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

and so μ satisfies $\overline{\mathrm{T}}^+(\frac{1}{2}\theta(a\,\cdot))$. By convexity of θ and since $\theta(0) = 0$, it holds $\frac{1}{2}\theta(2at) \geqslant \theta((a/2)t)$, and so μ satisfies $\overline{\mathrm{T}}^+(\theta((a/2)\,\cdot))$. The same reasoning yields to the conclusion that μ satisfies $\overline{\mathrm{T}}^-(\theta((a/2)\,\cdot))$, which completes the proof. □

Appendix : Proof of Theorem 5.1

The proof of Theorem 5.1 is a consequence of the following simple lemma of independent interest. This lemma is a new discrete Hardy type inequality. Hardy inequalities are commonly used to characterize measures that satisfy Poincaré inequality on the real line. We refer to [2] for more developments about the classical Hardy's inequality (see Theorem 6.2.1 in [2]).

Lemma 6.2. *Let h be a positive number and m be a positive measure on $\mathbb{R}$ invariant by translation of h. Let μ and ν be some non-negative measure on $\mathbb{R}$ with respective densities f and g with respect to m. Define, for all $y \in \mathbb{R}$,*

$$G_h(y) = \sum_{\ell=0}^{\infty} \frac{\mathbf{1}_{\ell h \leqslant y}}{g(y-\ell h)}, \quad \text{and} \quad F_h(y) = \sum_{\ell=0}^{\infty} f(y+\ell h),$$

with the convention $1/0 = +\infty$, and

$$B_h = \sup_{y \geqslant 0} G_h(y) F_h(y), \quad N_h = \sup_{y\geqslant 0} \frac{G_h(y-h)}{G_h(y)}, \quad M_h = \sup_{y \geqslant 0} \frac{F_h(y+h)}{F_h(y)},$$

where the supremum is the essential supremum with respect to the measure m. For all $\varphi : \mathbb{R} \to \mathbb{R}$ satisfying $\varphi(y) = 0$ for all $y < 0$, it holds

$$\int \varphi^2\, d\mu \leqslant B_h \left(1 + \sqrt{N_h}\right)\left(1 + \sqrt{M_h}\right) \int (\varphi(y) - \varphi(y-h))^2\, d\nu(y).$$

The proof of this lemma is postponed at the end of the Appendix.

In any case, we may observe that $M_h \leqslant 1$ and $N_h \leqslant 1$.

One typical example for which this result apply is when $\mu = \nu$ is the exponential probability measure on $\mathbb{R}^+$ denoted by τ^+ and defined by

$$\tau^+(dx) = e^{-x}\, dx$$

For simplicity, let us choose $h = 1$. In that case, one has

$$G_1(y) = \sum_{0 \leqslant \ell \leqslant y} e^{y-\ell} = \frac{e^{y+1} - e^{y-\lfloor y \rfloor}}{e-1}, \qquad F_1(y) = \sum_{\ell=0}^{\infty} e^{-y-\ell} = \frac{e^{-y}}{1-e^{-1}},$$

where $\lfloor y \rfloor$ denotes the integer part of y. It follows that

$$B_1 \leqslant \frac{e^2}{(e-1)^2}, \quad N_1 = M_1 = \frac{1}{e}.$$

And therefore Lemma 6.2 provides that for all function φ with $\varphi(y) = 0$ for $y < 0$,

$$\int \varphi^2 d\tau^+ \leqslant \frac{e}{(\sqrt{e}-1)^2} \int (\varphi(y) - \varphi(y-1))^2\, d\tau^+(y).$$

Let $f:\mathbb{R}\to\mathbb{R}$ be a function which is non-decreasing on $[-1,\infty)$ and such that $f(0)=0$; choosing $\varphi(y)=f(y)\mathbf{1}_{y\geqslant 0}$, and noting that $e/(\sqrt{e}-1)^2\leqslant 7$, one gets

$$\int f^2\,d\tau^+ \leqslant 7\int_0^1 f(y)^2 d\tau^+(y) + 7\int_1^\infty (f(y)-f(y-1))^2\,d\tau^+(y).$$

Since f is non-decreasing on $[-1,\infty)$ with $f(0)=0$, one has for all $y\in[0,1]$, $f(y)^2\leqslant (f(y)-f(y-1))^2$, and we obtain the following key lemma for the proof of Theorem 5.1.

Lemma 6.3. *For any function $f:\mathbb{R}\to\mathbb{R}$ non-decreasing on $[-1,\infty)$ and satisfying $f(0)=0$, one has*

$$\int f^2\,d\tau^+ \leqslant 7\int (f(y)-f(y-1))^2 d\tau^+(y).$$

We now turn to the proof of Theorem 5.1 which is an adaptation of the proof of the modified logarithmic Sobolev inequality for the exponential measure given in [7].

Proof of Theorem 5.1. We first prove (8), and then show that (9) is a simple consequence of (8).

The statement (8) is invariant by translation of the non-decreasing function f. Therefore, we may assume without loss of generality that $f(0)=0$. We also consider the non-decreasing function $\tilde f$ defined on $\mathbb{R}$ by

$$\tilde f(y) = -f(-y), \qquad y\in\mathbb{R}.$$

Since $u\log u\geqslant u-1$ for all $u\geqslant 0$, one has

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

where the last inequality comes from the fact the f and $\tilde f$ are non-negative on $\mathbb{R}^+$. Now suppose that the function f is such that $f(y)-f(y-1)\leqslant c$ for all $y\in\mathbb{R}$ and some $c>0$. The inequality (8) follows from the two next inequalities,

$$A=\int \tilde f^2\,d\tau^+ \leqslant 14e^{\sqrt{7}c+1}\int (f(y)-f(y-1))^2e^{f(y)}\,d\tau(y), \tag{17}$$

and if $c\leqslant\sqrt2/\sqrt7$ then

$$B=\int f^2e^f\,d\tau^+ \leqslant \frac{28}{2-7c^2}\int (f(y)-f(y-1))^2e^{f(y)}\,d\tau(y). \tag{18}$$

Thus we obtain (8) with the constant $K=\frac{7}{2-7c^2}+\frac72 e^{\sqrt7 c+1}$.

Let us prove (17). By Lemma 6.3 applied to the function $\tilde f$, one has

$$\begin{aligned}
A &\leqslant 7\int (\tilde f(y)-\tilde f(y-1))^2\,d\tau^+(y)\\
&\leqslant 7\exp\left(\frac{\int \tilde f(y)(\tilde f(y)-\tilde f(y-1))^2\,d\tau^+(y)}{\int (\tilde f(y)-\tilde f(y-1))^2\,d\tau^+(y)}\right)\\
&\qquad\qquad\int (\tilde f(y)-\tilde f(y-1))^2 e^{-\tilde f(y)}\,d\tau^+(y).
\end{aligned}$$

The last inequality is a consequence of the inequality $1 \leqslant \int e^{-\tilde f}\,d\nu\, e^{\int \tilde f\,d\nu}$ for the probability measure ν with density

$$\frac{d\nu}{d\tau^+}(y) = \frac{(\tilde f(y) - \tilde f(y-1))^2}{\int (\tilde f(y) - \tilde f(y-1))^2\,d\tau^+(y)}.$$

By Cauchy-Schwarz inequality and using Lemma 6.3 again, one has

$$\begin{aligned}
&\int \tilde f(y)(\tilde f(y) - \tilde f(y-1))^2\,d\tau^+(y)\\
&\qquad\leqslant \left(\int (\tilde f(y) - \tilde f(y-1))^4\,d\tau^+(y)\right)^{1/2}\left(\int \tilde f^2\,d\tau^+\right)^{1/2}\\
&\qquad\leqslant \sqrt{7}c\int (\tilde f(y) - \tilde f(y-1))^2\,d\tau^+(y).
\end{aligned}$$

It follows that

$$\begin{aligned}
A &\leqslant 7e^{\sqrt{7}c}\int (\tilde f(y) - \tilde f(y-1))^2 e^{-\tilde f(y)}\,d\tau^+(y)\\
&= 7e^{\sqrt{7}c}\int_{-\infty}^0 (f(y+1) - f(y))^2 e^{f(y)} e^y\,dy\\
&= 7e^{\sqrt{7}c-1}\int_{-\infty}^1 (f(y) - f(y-1))^2 e^{f(y-1)} e^y\,dy\\
&\leqslant 14e^{\sqrt{7}c+1}\int_{-\infty}^1 (f(y) - f(y-1))^2 e^{f(y)}\,d\tau(y).
\end{aligned}$$

Now, let us prove (18). To that purpose, we want to apply Lemma 6.3 to the function $g = fe^{f/2}$. First let us show that g is non-decreasing on $[-1,\infty)$. Since the function $x \mapsto xe^{x/2}$ is non-increasing on $[-2,\infty)$ and f is non-decreasing on $\mathbb{R}$, it is enough to check that $f \geqslant -2$ on $[-1,\infty)$. But, $f(-1) \geqslant f(0) - c = -c \geqslant -1$ since by assumption $c \leqslant 1$. Applying Lemma 6.3 to the function $fe^{f/2}$ and using the inequality

$$0 \leqslant be^{b/2} - ae^{a/2} \leqslant (b-a)(1+\frac{b}{2})e^{b/2}, \quad -2 \leqslant a \leqslant b,$$

one gets

$$\begin{aligned}
B &\leqslant 7\int \left(f(y)e^{f(y)/2} - f(y-1)e^{f(y-1)}\right)^2\,d\tau^+(y)\\
&\leqslant 14\int (f(y) - f(y-1))^2 e^{f(y)}\,d\tau^+(y)\\
&\qquad\qquad + \frac{7}{2}\int f^2(y)(f(y) - f(y-1))^2 e^{f(y)}\,d\tau^+(y)\\
&\leqslant 14\int (f(y) - f(y-1))^2 e^{f(y)}\,d\tau^+(y) + \frac{7c^2}{2}B.
\end{aligned}$$

This provides inequality (18) when $c \leqslant \sqrt{2}/\sqrt{7}$. The proof of the first inequality (8) of Theorem 5.1 is completed.

To obtain the second inequality (9) of Theorem 5.1 from (8), it suffices to observe that by a simple change of variables

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

This ends the proof of Theorem 5.1. □

Proof of Lemma 6.2. Since $\varphi(y) = 0$ for all $y < 0$, one has for all $y \geqslant 0$

$$\varphi(y) = \sum_{k \in \mathbb{N}, hk \leqslant y} (\varphi(y - kh) - \varphi(y - (k+1)h)).$$

By Cauchy-Schwarz inequality, it follows that

$$\varphi^2(y) \leqslant \left(\sum_{k=0}^{\infty} (\varphi(y - kh) - \varphi(y - (k+1)h))^2 g(y - kh) \sqrt{G_h(y - kh)} \right) \left(\sum_{k=0}^{\infty} \frac{\mathbf{1}_{hk \leqslant y}}{g(y - hk)\sqrt{G_h(y - kh)}} \right).$$

As a consequence, by Fubini's theorem, by the translation invariant property of the measure m, and since $H(y) = 0$ for $y < 0$, one has

$$\begin{aligned}
\int \varphi^2 \, d\mu &\leqslant \sum_{k=0}^{\infty} \int_{\mathbb{R}^+} (\varphi(y - kh) - \varphi(y - (k+1)h))^2 g(y - kh) \sqrt{G_h(y - kh)} \\
&\qquad \left(\sum_{\ell=0}^{\infty} \frac{\mathbf{1}_{\ell h \leqslant y}}{g(y - h\ell)\sqrt{G_h(y - \ell h)}} \right) f(y) \, dm(y) \\
&= \sum_{k=0}^{\infty} \int_{\mathbb{R}^+} (\varphi(y) - \varphi(y - h))^2 g(y) \sqrt{G_h(y)} \\
&\qquad \left(\sum_{\ell=0}^{\infty} \frac{\mathbf{1}_{h\ell \leqslant y + kh}}{g(y + kh - \ell h)\sqrt{G_h(y + kh - \ell h)}} \right) f(y + kh) \, dm(y) \\
&= \int_{\mathbb{R}^+} (\varphi(y) - \varphi(y - h))^2 g(y) \sqrt{G_h(y)} C_h(y) \, dm(y),
\end{aligned}$$

where $C_h(y) = \sum_{k=0}^{\infty} \left(\sum_{\ell=0}^{\infty} \frac{\mathbf{1}_{h\ell \leqslant y + kh}}{g(y + kh - \ell h)\sqrt{G_h(y + kh - \ell h)}} \right) f(y + kh)$.
To end the proof of Lemma 6.2, it remains to show that for m-almost every $y \geqslant 0$,

$$\sqrt{G_h(y)} C_h(y) \leqslant (1 + \sqrt{N_h}) B_h (1 + \sqrt{M_h}).$$

Observe that $G_h = 0$ on $(-\infty, 0)$ and that

$$\frac{1}{g(y + kh - \ell h)} = G_h(y + kh - \ell h) - G_h(y + kh - (\ell + 1)h), \qquad h\ell \leqslant y + kh,$$

and

$$f(y + kh) = F_h(y + kh) - F_h(y + (k+1)h), \qquad y + kh \geqslant 0.$$

Therefore, using the definitions of the constants M_h, N_h and B_h, we get for all $y \geqslant 0$

$$\begin{aligned}
C_h(y) &= \sum_{k=0}^{\infty} \left(\sum_{\ell, h\ell \leqslant y+kh} \frac{G_h(y+kh-\ell h) - G_h(y+kh-(\ell+1)h)}{\sqrt{G_h(y+kh-\ell h)}} \right) f(y+kh) \\
&\leqslant \sum_{k=0}^{\infty} \left(\sum_{\ell, h\ell \leqslant y+kh} \left(\sqrt{G_h(y+kh-\ell h)} - \sqrt{G_h(y+kh-(\ell+1)h)} \right) (1+\sqrt{N_h}) \right) \\
&\qquad \cdot f(y+kh) \\
&= (1+\sqrt{N_h}) \sum_{k=0}^{\infty} \sqrt{G_h(y+kh)} f(y+kh) \\
&\leq (1+\sqrt{N_h}) \sqrt{B_h} \sum_{k=0}^{\infty} \frac{f(y+kh)}{\sqrt{F_h(y+kh)}} \\
&= (1+\sqrt{N_h}) \sqrt{B_h} \sum_{k=0}^{\infty} \frac{F_h(y+kh) - F_h(y+(k+1)h)}{\sqrt{F_h(y+kh)}} \\
&\leqslant (1+\sqrt{N_h}) \sqrt{B_h} (1+\sqrt{M_h}) \sum_{k=0}^{\infty} \left(\sqrt{F_h(y+kh)} - \sqrt{F_h(y+(k+1)h)} \right) \\
&= (1+\sqrt{N_h}) \sqrt{B_h} (1+\sqrt{M_h}) \sqrt{F_h(y)} \\
&\leqslant (1+\sqrt{N_h}) B_h (1+\sqrt{M_h}) \frac{1}{\sqrt{G_h(y)}}
\end{aligned}$$

The last equality above holds since $F_h(y+kh) \to 0$ when $k \to \infty$. The proof of Lemma 6.2 is completed. □

References

1. Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below*, Invent. Math. **195** (2014), no. 2, 289–391. MR-3152751 20
2. Cécile Ané, Sébastien Blachère, Djalil Chafaï, Pierre Fougères, Ivan Gentil, Florent Malrieu, Cyril Roberto, and Grégory Scheffer, *Sur les inégalités de Sobolev logarithmiques*, Panoramas et Synthèses [Panoramas and Syntheses], vol. 10, Société Mathématique de France, Paris, 2000, With a preface by Dominique Bakry and Michel Ledoux. MR-1845806 23
3. Anders Björner, Michel Las Vergnas, Bernd Sturmfels, Neil White, and Günter M. Ziegler, *Oriented matroids*, second ed., Encyclopedia of Mathematics and its Applications, vol. 46, Cambridge University Press, Cambridge, 1999. MR-1744046 10
4. Sergey G. Bobkov, Ivan Gentil, and Michel Ledoux, *Hypercontractivity of Hamilton-Jacobi equations*, J. Math. Pures Appl. (9) **80** (2001), no. 7, 669–696. MR-1846020 6, 17, 19, 20
5. Sergey G. Bobkov and Friedrich Götze, *Discrete isoperimetric and Poincaré-type inequalities*, Probab. Theory Related Fields **114** (1999), no. 2, 245–277. MR-1701522 5, 17
6. ______, *Exponential integrability and transportation cost related to logarithmic Sobolev inequalities*, J. Funct. Anal. **163** (1999), no. 1, 1–28. MR-1682772 17
7. Sergey G. Bobkov and Michel Ledoux, *Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution*, Probab. Theory Related Fields **107** (1997), no. 3, 383–400. MR-1440138 18, 24
8. Stéphane Boucheron, Gábor Lugosi, and Pascal Massart, *Concentration inequalities*, Oxford University Press, Oxford, 2013, A nonasymptotic theory of independence, With a foreword by Michel Ledoux. MR-3185193 1, 2
9. Stamatis Cambanis, Gordon Simons, and William Stout, *Inequalities for $Ek(X,Y)$ when the marginals are fixed*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete **36** (1976), no. 4, 285–294. MR-0420778 5
10. Giorgio Dall'Aglio, *Sugli estremi dei momenti delle funzioni di ripartizione doppia*, Ann. Scuoloa Norm. Sup. Pisa (3) **10** (1956), 35–74. MR-0081577 5

11. Hacène Djellout, Arnaud Guillin, and Liming Wu, *Transportation cost-information inequalities and applications to random dynamical systems and diffusions*, Ann. Probab. **32** (2004), no. 3B, 2702–2732. MR-2078555 2
12. Naomi Feldheim, Arnaud Marsiglietti, Piotr Nayar, and Jing Wang, *A note on the convex infimum convolution inequality*, Preprint arXiv:1505.00240, 2015. 6
13. Maurice Fréchet, *Sur les tableaux dont les marges et des bornes sont données*, Rev. Inst. Internat. Statist. **28** (1960), 10–32. MR-0115232 5
14. Nathael Gozlan, *A characterization of dimension free concentration in terms of transportation inequalities*, Ann. Probab. **37** (2009), no. 6, 2480–2498. MR-2573565 2
15. ______, *Transport-entropy inequalities on the line*, Electron. J. Probab. **17** (2012), no. 49, 18. MR-2946156 2, 3, 4, 5, 22
16. Nathael Gozlan and Christian Léonard, *Transport inequalities. A survey*, Markov Process. Related Fields **16** (2010), no. 4, 635–736. MR-2895086 1, 2, 17
17. Nathael Gozlan, Cyril Roberto, and Paul-Marie Samson, *Hamilton Jacobi equations on metric spaces and transport entropy inequalities*, Rev. Mat. Iberoam. **30** (2014), no. 1, 133–163. MR-3186934 6, 19, 20
18. Nathael Gozlan, Cyril Roberto, Paul-Marie Samson, and Prasad Tetali, *Kantorovich duality for general transport costs and applications*, Preprint hal-01098114v2, 2015. 3, 4, 6, 8, 16, 17, 19
19. Francis Hirsch, Christophe Profeta, Bernard Roynette, and Marc Yor, *Peacocks and associated martingales, with explicit constructions*, Bocconi & Springer Series, vol. 3, Springer, Milan; Bocconi University Press, Milan, 2011. MR-2808243 7
20. Wassily Hoeffding, *Maßstabinvariante korrelationstheorie.*, Schriften des Mathematischen Instituts und des Instituts für Angewandte Mathematik der Universität Berlin **5** (1940), 181–233. 5
21. Michel Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001. MR-1849347 1, 2
22. Albert W. Marshall, Ingram Olkin, and Barry C. Arnold, *Inequalities: theory of majorization and its applications*, second ed., Springer Series in Statistics, Springer, New York, 2011. MR-2759813 7
23. Katalin Marton, *A simple proof of the blowing-up lemma*, IEEE Trans. Inform. Theory **32** (1986), no. 3, 445–446. MR-838213 1
24. ______, *Bounding $\bar{d}$-distance by informational divergence: a method to prove measure concentration*, Ann. Probab. **24** (1996), no. 2, 857–866. MR-1404531 1
25. ______, *A measure concentration inequality for contracting Markov chains*, Geom. Funct. Anal. **6** (1996), no. 3, 556–571. MR-1392329 1
26. ______, *An inequality for relative entropy and logarithmic Sobolev inequalities in Euclidean spaces*, J. Funct. Anal. **264** (2013), no. 1, 34–61. MR-2995699 2
27. Bernard Maurey, *Some deviation inequalities*, Geom. Funct. Anal. **1** (1991), no. 2, 188–197. MR-1097258 3
28. Richard Rado, *An inequality*, J. London Math. Soc. **27** (1952), 1–6. MR-0045168 5, 9
29. Paul-Marie Samson, *Concentration inequalities for convex functions on product spaces*, Stochastic inequalities and applications, Progr. Probab., vol. 56, Birkhäuser, Basel, 2003, pp. 33–52. MR-2073425 3
30. Volker Strassen, *The existence of probability measures with given marginals*, Ann. Math. Statist. **36** (1965), 423–439. MR-0177430 5, 7
31. Michel Talagrand, *Transportation cost for Gaussian and other product measures*, Geom. Funct. Anal. **6** (1996), no. 3, 587–600. MR-1392331 1
32. Cédric Villani, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009, Old and new. MR-2459454 1, 15, 17
33. Neng-Yi Wang, *Concentration inequalities for Gibbs sampling under d_{l_2}-metric*, Electron. Commun. Probab. **19** (2014), no. 63, 11. MR-3262069 2

Université Paris Est Marne la Vallée - Laboratoire d'Analyse et de Mathématiques Appliquées (UMR CNRS 8050), 5 bd Descartes, 77454 Marne la Vallée Cedex 2, France

Université Paris Ouest Nanterre La Défense - Modal'X, 200 avenue de la République 92000 Nanterre, France

School of Mathematics & School of Computer Science, Georgia Institute of Technology, Atlanta, GA 30332

E-mail address: nathael.gozlan@univ-mlv.fr, cyril.roberto@math.cnrs.fr, paul-marie.samson@univ-mlv.fr, yan.shu@ens-lyon.fr, tetali@math.gatech.edu