Elect. Comm. in Probab. **11** (2006), 64–77

ELECTRONIC COMMUNICATIONS in PROBABILITY

INTEGRAL CRITERIA FOR TRANSPORTATION-COST INEQUALITIES

NATHAEL GOZLAN
Équipe Modal-X, Université Paris 10. 200 av. de la République. 92001 Nanterre Cedex, France
email: nathael.gozlan@u-paris10.fr

Submitted 3 December 2005, *accepted in final form* 5 June 2006

AMS 2000 Subject classification: 60E15 and 46E30
Keywords: Transportation-cost inequalities and Orlicz Spaces

Abstract
In this paper, we provide a characterization of a large class of transportation-cost inequalities in terms of exponential integrability of the cost function under the reference probability measure. Our results completely extend the previous works by Djellout, Guillin and Wu [8] and Bolley and Villani [3].

1 Introduction

In all the paper, $(\mathcal{X}, d)$ will be a polish space equipped with its Borel σ-field. The set of probability measures on $\mathcal{X}$ will be denoted by $\mathcal{P}(\mathcal{X})$.

1.1 Norm-entropy inequalities and transportation cost inequalities

The aim of this paper is to give necessary and sufficient conditions for inequalities of the following form :

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \alpha\left(\|\nu - \mu\|_\Phi^*\right) \leq \mathrm{H}(\nu \mid \mu), \tag{1.1}$$

where

- $\alpha : \mathbb{R}^+ \to \mathbb{R}^+ \cup \{+\infty\}$ is a convex lower semi-continuous function vanishing at 0,
- The semi-norm $\|\nu - \mu\|_\Phi^*$ is defined by

$$\|\nu - \mu\|_\Phi^* := \sup_{\varphi \in \Phi} \left\{ \int_{\mathcal{X}} \varphi \, d\nu - \int_{\mathcal{X}} \varphi \, d\mu \right\}, \tag{1.2}$$

  where Φ is a set of bounded measurable functions on $\mathcal{X}$ which is symmetric, i.e.

$$\varphi \in \Phi \Rightarrow -\varphi \in \Phi,$$

- The quantity $\mathrm{H}(\nu \mid \mu)$ is the relative entropy of ν with respect to μ defined by

$$\mathrm{H}(\nu \mid \mu) = \int_{\mathcal{X}} \log \frac{d\nu}{d\mu}\, d\nu,$$

  if ν is absolutely continuous with respect to μ and $+\infty$ otherwise.

Inequalities of the form (1.1) were introduced by C. Léonard and the author in [12]. They are called *norm-entropy inequalities*. An important particular case, is when Φ is the set of all bounded 1-Lipschitz functions on $\mathcal{X}$: $\Phi = \mathrm{BLip}_1(\mathcal{X}, d)$. Indeed, in that case $\|\nu - \mu\|_\Phi^*$ is the optimal transportation cost between ν and μ associated to the metric cost function $d(x, y)$. Let us recall that if $c : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$ is a lower semi-continuous function, then the optimal transportation cost between $\nu \in \mathcal{P}(\mathcal{X})$ and $\mu \in \mathcal{P}(\mathcal{X})$ is defined by

$$\mathcal{T}_c(\nu, \mu) = \inf \int_{\mathcal{X}^2} c(x, y)\, d\pi(x, y) \tag{1.3}$$

where π describes the set $\Pi(\nu, \mu)$ of all probability measures on $\mathcal{X} \times \mathcal{X}$ having ν for first marginal and μ for second marginal. According to Kantorovich-Rubinstein duality theorem (see e.g Theorem 1.3 of [18]), if the cost function c is the metric d, the following identity holds

$$\mathcal{T}_d(\nu, \mu) = \sup_{\varphi \in \mathrm{BLip}_1(\mathcal{X}, d)} \left\{ \int_{\mathcal{X}} \varphi\, d\nu - \int_{\mathcal{X}} \varphi\, d\mu \right\}. \tag{1.4}$$

In this setting, inequality (1.1) becomes

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \alpha\left(\mathcal{T}_d(\nu, \mu)\right) \leq \mathrm{H}(\nu \mid \mu) \tag{1.5}$$

Such an inequality is called a *convex transportation-cost inequality* (convex T.C.I).

1.2 Applications of transportation-cost inequalities

After the seminal works of K. Marton [14, 15] and M. Talagrand [17], new efforts have been made in order to understand this kind of inequalities. The reason of this interest is the link between T.C.I and concentration of measure inequalities. Namely, according to a general argument due to K. Marton, if μ satisfies (1.5), then μ has the following concentration property

$$\forall A \subset \mathcal{X} \text{ s.t. } \mu(A) \geq \frac{1}{2}, \quad \forall \varepsilon \geq r, \quad \mu(A^\varepsilon) \geq 1 - e^{-\alpha(\varepsilon - r)},$$

with $r = \alpha^{-1}(\log(2))$ and $A^\varepsilon = \{x \in \mathcal{X} : d(x, A) \leq \varepsilon\}$. For a proof of this fact, see e.g. Theorem 9 of [12]. Other applications of T.C.Is were investigated in [8], [3], [2] and [12]. In these papers, it was shown that T.C.Is are an efficient way for deriving precise deviations results for Markov chains and empirical processes. One can also consult [5] and [10] for applications of norm-entropy inequalities to the study of conditional principles of Gibbs type for empirical measures and random weighted measures.

1.3 Necessary and sufficient conditions for norm-entropy inequalities

Our main result gives necessary and sufficient conditions on μ for (1.1) to be satisfied. Before to state it, let us introduce some notations. In all what follows, $\mathcal{C}$ will denote the set of convex functions $\alpha : \mathbb{R}^+ \to \mathbb{R}^+ \cup \{+\infty\}$ which are lower semi continuous and such that $\alpha(0) = 0$. For a given α, the monotone convex conjugate of α will be denoted by $\alpha^{\circledast}$. It is defined by

$$\forall s \geq 0, \quad \alpha^{\circledast}(s) = \sup_{t\geq 0} \{st - \alpha(t)\}.$$

Note that, if α belongs to $\mathcal{C}$, then $\alpha^{\circledast}$ also belongs to $\mathcal{C}$. Furthermore, one has the relation $\alpha^{\circledast\circledast} = \alpha$. If α is in $\mathcal{C}$, the Orlicz space $\mathbb{L}_{\tau_\alpha}(\mathcal{X}, \mu)$ associated to the function $\tau_\alpha := e^\alpha - 1$ is defined by

$$\mathbb{L}_{\tau_\alpha}(\mathcal{X}, \mu) = \left\{ f : \mathcal{X} \to \mathbb{R} \text{ such that } \exists \lambda > 0, \int_{\mathcal{X}} \tau_\alpha\left(\frac{f}{\lambda}\right) d\mu < +\infty \right\},$$

where μ almost everywhere equal functions are identified. The space $\mathbb{L}_{\tau_\alpha}(\mathcal{X}, \mu)$ is equipped with its classical Luxemburg norm $\| \,.\, \|_{\tau_\alpha}$, i.e

$$\forall f \in \mathbb{L}_{\tau_\alpha}(\mathcal{X}, \mu), \quad \|f\|_{\tau_\alpha} = \inf\left\{ \lambda > 0 \text{ such that } \int_{\mathcal{X}} \tau_\alpha\left(\frac{f}{\lambda}\right) d\mu \leq 1 \right\}.$$

We will need the following assumptions on α :

Assumptions.
(A_1) *The effective domain of $\alpha^{\circledast}$ is open on the right, i.e $\{s \in \mathbb{R}^+ : \alpha^{\circledast}(s) < +\infty\} = [0, b[$, for some $b > 0$.*
(A_2) *The function $\alpha^{\circledast}$ is super-quadratic near 0, i.e*

$$\exists s_{\alpha^\circledast} > 0, c_{\alpha^\circledast} > 0, \quad \forall s \in [0, s_{\alpha^\circledast}], \quad \alpha^{\circledast}(s) \geq c_{\alpha^\circledast} s^2. \tag{1.6}$$

We can now state the main result of this paper, which will be proved in section 2.

Theorem 1.7. *Let $\alpha \in \mathcal{C}$ satisfy assumptions (A_1) and (A_2) and $\mu \in \mathcal{P}(\mathcal{X})$. The following statements are equivalent :*

1. $\exists a > 0$ *such that* , $\quad \forall \nu \in \mathcal{P}(\mathcal{X}), \quad \alpha\left(\dfrac{\|\nu - \mu\|^*_\Phi}{a}\right) \leq \mathrm{H}(\nu \mid \mu)$
2. $\exists M > 0$ *such that* , $\quad \forall \varphi \in \Phi, \quad \|\varphi - \langle \varphi, \mu \rangle\|_{\tau_\alpha} \leq M.$

More precisely, if (1) holds true then one can take $M = 3a$. Conversely, if (2) holds true, then one can take $a = \sqrt{2} m_\alpha M$, with m_α defined by
$m_\alpha = e \min\left\{ \max\left(\frac{1}{\alpha^{-1}(2)\sqrt{c_{\alpha^\circledast}(1-u)}}, \frac{1}{u} \right) : u \in]0,1[\text{ such that } \frac{u}{\sqrt{1-u}} \leq s_{\alpha^\circledast}\sqrt{c^\circledast_\alpha} \text{ and } \frac{u^3}{1-u} \leq 2 \right\},$
where the constants $s_{\alpha^\circledast}$ and $c_{\alpha^\circledast}$ are given by (1.6).

Remark 1.8.

- *If Φ contains an element which is not μ-a.e constant, and if inequality (1.1) holds for some $\alpha \in \mathcal{C}$, then α satisfies assumption A_2 (see Lemma 2.1).*
- *The constant $a = \sqrt{2}m_\alpha M$ is not optimal. This can be easily checked by considering the celebrated Pinsker inequality, i.e*

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \frac{\|\nu - \mu\|_{TV}^2}{2} \leq \mathrm{H}(\nu \mid \mu), \tag{1.9}$$

where $\|\nu - \mu\|_{TV}$ is the total-variation norm which is defined by

$$\|\nu - \mu\|_{TV} = \sup\left\{ \int_{\mathcal{X}} \varphi \, d\nu - \int_{\mathcal{X}} \varphi \, d\mu, |\varphi| \leq 1 \right\}.$$

In this example, $\alpha(x) = x^2$ and the optimal constant is $a_0 = \sqrt{2}$. On the other hand, Theorem 1.7 yields the constant $a_1 = \sqrt{2}m_{x^2}M$, with $M = \sup_{|\varphi|\leq 1} \|\varphi - \langle \varphi, \mu\rangle\|_{\tau_{x^2}}$. It is easy to check that $m_{x^2} = 2e$ and that $\frac{1}{2\sqrt{\log(2)}} \leq M \leq \frac{2}{\sqrt{\log(2)}}$, thus $\frac{e}{\sqrt{\log(2)}} \leq \frac{a_1}{a_0} \leq \frac{4e}{\sqrt{\log(2)}}$.

In order to prove Theorem 1.7, we will take advantage of the dual formulation of norm-entropy inequalities developed in [12]. Namely, according to Theorem 3.15 of [12], we have the following result :

Theorem 1.10. *The inequality*

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \alpha\left(\frac{\|\nu - \mu\|_\Phi^*}{a}\right) \leq \mathrm{H}(\nu \mid \mu),$$

with $\alpha \in \mathcal{C}$ is equivalent to the following condition :

$$\forall \varphi \in \Phi, \quad \forall s \in \mathbb{R}^+, \quad \int_{\mathcal{X}} e^{s\varphi} \, d\mu \leq e^{s\langle \varphi, \mu\rangle + \alpha^{\circledast}(as)}. \tag{1.11}$$

According to (1.11), the only thing to know is how to majorize the Laplace transform of a centered random variable X knowing that this random variable satisfies an Orlicz integrability condition of the form : $\mathbb{E}\left[e^{\alpha\left(\frac{X}{\lambda}\right)}\right] < +\infty$, for some $\lambda > 0$. Estimates of this kind are very useful in probability theory, because they enable us to control the deviation probabilities of sums of independent and identically distributed random variables. In [12], we have shown how to deduce Pinsker inequality from the classical Hoeffding estimate (see Section 2.3 of [12]). We also proved that the weighted version of Pinsker inequality (1.21) recently obtained by Bolley and Villani in [3] is a consequence of Bernstein estimate (see Corollaries 3.23 and 3.24 of [12]). Here, Theorem 1.7 will follow very easily from the following theorem which is due to Kozachenko and Ostrovskii (see [13] and [4] p. 63-68) :

Theorem 1.12. *Suppose that $\alpha \in \mathcal{C}$ satisfies Assumptions (A_1) and (A_2), then for all $f \in \mathbb{L}_{\tau_\alpha}(\mathcal{X}, \mu)$ such that $\int_{\mathcal{X}} f \, d\mu = 0$, the following holds*

$$\forall s \geq 0, \quad \int_{\mathcal{X}} e^{sf} \, d\mu \leq e^{\alpha^{\circledast}(as)},$$

with $a = \sqrt{2}m_\alpha \|f\|_{\tau_\alpha}$, where m_α is the constant defined in Theorem 1.7.

For further informations on the preceding result, we refer to Chapter VII of [11] (p. 193-197) where a complete detailed proof is given. Before proving Theorem 1.7, we discuss below some of its applications.

1.4 Applications to T.C.Is

Applying the preceding theorem to the case where Φ is the Lipschitz ball $\mathrm{BLip}_1(\mathcal{X}, d)$, one obtains the following result.

Theorem 1.13. *Let $\alpha \in \mathcal{C}$ satisfy assumptions (A_1) and (A_2) and $\mu \in \mathcal{P}(\mathcal{X})$ be such that $\int_{\mathcal{X}} d(x_0, x)\, d\mu(x) < +\infty$ for all $x_0 \in \mathcal{X}$. The following statements are equivalent :*

1. *$\exists a > 0$ such that $\forall \nu \in \mathcal{P}(\mathcal{X})$,* $\quad \alpha\left(\dfrac{\mathcal{T}_d(\nu, \mu)}{a}\right) \le \mathrm{H}(\nu \mid \mu).$
2. *For all $x_0 \in \mathcal{X}$, the function $d(x_0, \,.\,) \in \mathbb{L}_{\tau_\alpha}(\mathcal{X}, \mu)$.*

More precisely, if (2) holds true, then one can take $a = 2\sqrt{2} m_\alpha \inf_{x_0 \in \mathcal{X}} \|d(x_0, \,.\,)\|_{\tau_\alpha}$, where m_α was defined in Theorem 1.7.

Remark 1.14. *In other words, μ satisfies the transportation-cost inequality (1) if and only if there is some $\delta > 0$ such that $\int_{\mathcal{X}} e^{\alpha(\delta d(x_0, x))}\, d\mu(x) < +\infty$, for some (equivalently, for all) $x_0 \in \mathcal{X}$.*

Actually, other transportation cost inequalities can be deduced from Theorem 1.7. Using a majorization technique developed by F. Bolley and C. Villani in [3], we will prove the following result :

Theorem 1.15. *Let $c(\,.\,,\,.\,)$ be a cost function such that $c(x, y) = q(d(x, y))$, where $q : \mathbb{R}^+ \to \mathbb{R}^+$ is an increasing convex function satisfying the Δ_2-condition, i.e*

$$\exists K > 0, \quad \forall x \in \mathbb{R}^+, \quad q(2x) \le K q(x). \tag{1.16}$$

If $\alpha \in \mathcal{C}$ satisfies assumptions (A_1) and (A_2), then for all $\mu \in \mathcal{P}(\mathcal{X})$ such that $\int_{\mathcal{X}} c(x_0, x)\, d\mu(x) < +\infty$ for all $x_0 \in \mathcal{X}$, the following statements are equivalent :

1. *$\exists a > 0$,* $\quad \forall \nu \in \mathcal{P}(\mathcal{X}), \quad \alpha\left(\dfrac{\mathcal{T}_c(\nu, \mu)}{a}\right) \le \mathrm{H}(\nu \mid \mu),$
2. *For all $x_0 \in \mathcal{X}$, the function $c(x_0, \,.\,) \in \mathbb{L}_{\tau_\alpha}(\mathcal{X}, \mu)$.*

More precisely, if (2) holds true then one can take $a = \sqrt{2} K m_\alpha \inf_{x_0 \in \mathcal{X}} \|c(x_0, \,.\,)\|_{\tau_\alpha}$. Furthermore, if $\operatorname{dom} \alpha = \mathbb{R}^+$ then the following inequality holds

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \mathcal{T}_c(\nu, \mu) \le \sqrt{2} K m_\alpha \inf_{x_0 \in \mathcal{X},\, \delta > 0} \frac{1}{\delta} \left(1 + \frac{\log \int_{\mathcal{X}} e^{\delta \alpha(c(x_0, x))}\, d\mu(x)}{\log 2}\right) \alpha^{-1}\left(\mathrm{H}(\nu \mid \mu)\right) \tag{1.17}$$

Contrary to what happens in the case where c is the metric d, a transportation-cost inequality $\alpha\left(\mathcal{T}_c(\nu, \mu)\right) \le \mathrm{H}(\nu \mid \mu)$ can hold even if α does not satisfy Assumption (A_2). The most known example is Talagrand inequality, also called $\mathbb{T}_2$-inequality. Let us recall that a probability measure μ on $\mathbb{R}^n$ satisfies the Talagrand inequality $\mathbb{T}_2(a)$ if

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \mathcal{T}_{d^2}(\nu, \mu) \le a\, \mathrm{H}(\nu \mid \mu), \tag{1.18}$$

where $d(x,y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$. Gaussian measures do satisfy a $\mathbb{T}_2$-inequality. This was first shown by Talagrand in [17]. In this case, the corresponding α is a linear function and hence its monotone conjugate $\alpha^{\circledast}$ does not satisfy (A_2). Sufficient conditions are known for Talagrand inequality. In [16], it was shown by F. Otto and C. Villani that if $d\mu = e^{-\Phi} dx$ is a probability measure on $\mathbb{R}^n$ satisfying a logarithmic Sobolev inequality with constant a, then it also satisfies the inequality $\mathbb{T}_2(a)$. Furthermore, if μ satisfies $\mathbb{T}_2(a)$, then it satisfies the Poincaré inequality with a constant $a/2$. An alternative proof of these facts was proposed in [1] by S.G. Bobkov, I. Gentil and M. Ledoux. In a recent paper P. Cattiaux and A. Guillin gave an example of a probability measure satisfying $\mathbb{T}_2$ but not the logarithmic Sobolev inequality (see [6]). A necessary and sufficient condition for $\mathbb{T}_2$ is not yet known. Other examples of transportation-cost inequalities involving a linear α can be found in [1], [9] and [6]. The common feature of these $\mathbb{T}_2$-like inequalities is that they enjoy a dimension free tensorization property (see e.g Theorem 4.12 of [12]) which in turn implies a dimension free concentration phenomenon.

1.5 About the literature

Theorems 1.15 and 1.13 extend previous results obtained by H. Djellout, A. Guillin and L. Wu in [8] and by F. Bolley and C. Villani in [3].
In [8], H. Djellout, A. Guillin and L. Wu obtained the first integral criteria for the so called $\mathbb{T}_1$-inequality. Let us recall that a probability measure μ on $\mathcal{X}$ is said to satisfy the inequality $\mathbb{T}_1(a)$ if

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \mathcal{T}_d(\nu,\mu)^2 \leq a\,\mathrm{H}(\nu \mid \mu). \tag{1.19}$$

According to Jensen inequality, $\mathcal{T}_d(\nu,\mu)^2 \leq \mathcal{T}_{d^2}(\nu,\mu)$, and thus $\mathbb{T}_2(a) \Rightarrow \mathbb{T}_1(a)$. The inequality $\mathbb{T}_1$ is weaker than $\mathbb{T}_2$ and it is also considerably easier to study. According to Theorem 3.1 of [8], the following propositions are equivalent :

1. $\exists a > 0,$ such that μ satisfies $\mathbb{T}_1(a)$
2. $\exists \delta > 0$ such that $\displaystyle\int_{\mathcal{X}^2} e^{\delta d(x,y)^2}\, d\mu(x)d\mu(y) < +\infty$

More precisely, if $\displaystyle\int_{\mathcal{X}^2} e^{\delta d(x,y)^2}\, d\mu(x)d\mu(y) < +\infty$ for some $\delta > 0$, then one can take

$$a = \frac{4}{\delta^2} \sup_{k\geq 1} \left(\frac{(k!)^2}{(2k!)} \right)^{1/k} \left[\int_{\mathcal{X}^2} e^{\delta^2 d(x,y)^2}\, d\mu(x)d\mu(y) \right]^{1/k} < +\infty. \tag{1.20}$$

The link between the constants a and δ was then improved by F. Bolley and C. Villani in [3] (see (1.25) bellow).

In [3], F. Bolley and C. Villani obtained the following weighted versions of Pinsker inequality : if $\chi : \mathcal{X} \to \mathbb{R}^+$, is a measurable function, then for all $\nu \in \mathcal{P}(\mathcal{X})$,

$$\|\chi \cdot (\nu - \mu)\|_{TV} \leq \left(\frac{3}{2} + \log \int_{\mathcal{X}} e^{2\chi}\, d\mu \right) \left(\sqrt{\mathrm{H}(\nu \mid \mu)} + \frac{1}{2}\,\mathrm{H}(\nu \mid \mu) \right) \tag{1.21}$$

$$\|\chi \cdot (\nu - \mu)\|_{TV} \leq \sqrt{1 + \log \int_{\mathcal{X}} e^{\chi^2}\, d\mu}\sqrt{2\,\mathrm{H}(\nu \mid \mu)} \tag{1.22}$$

Using the following upper bound (see [18], prop. 7.10)

$$\mathcal{T}_{d^p}(\nu,\mu) \leq 2^{p-1}\|d(x_0,\,.\,)^p \cdot (\nu-\mu)\|_{TV}, \tag{1.23}$$

they deduce from (1.21) and (1.22) the following transportation cost inequalities involving cost functions of the form $c(x,y)=d(x,y)^p$ with $p\geq 1$: $\forall \nu \in \mathcal{P}(\mathcal{X})$,

$$\mathcal{T}_{d^p}(\nu,\mu)^{1/p} \leq 2 \inf_{x_0\in\mathcal{X},\,\delta>0}\left[\frac{1}{\delta}\left(\frac{3}{2}+\log\int_{\mathcal{X}} e^{\delta d(x_0,x)^p}\,d\mu(x)\right)\right]^{1/p}\cdot\left[\mathrm{H}(\nu\mid\mu)^{1/p}+\left(\frac{\mathrm{H}(\nu\mid\mu)}{2}\right)^{1/2p}\right], \tag{1.24}$$

$$\mathcal{T}_{d^p}(\nu,\mu) \leq 2 \inf_{x_0\in\mathcal{X},\,\delta>0}\left[\frac{1}{2\delta}\left(1+\log\int_{\mathcal{X}} e^{\delta d(x_0,x)^{2p}}\,d\mu(x)\right)\right]^{1/2p}\cdot \mathrm{H}(\nu\mid\mu)^{1/2p}. \tag{1.25}$$

Note that for $p=1$, the constant in (1.25) is sharper than (1.20). Note also that, up to numerical factors, (1.24) and (1.25) are particular cases of (1.17).

In order to derive T.C.Is from norm-entropy inequalities, we will follow the lines of [3]. To do this, we will deduce from Theorem 1.7 a general version of weighted Pinsker inequality (see Theorem 2.7). Theorem 1.15 will follow from Theorem 2.7 and from Lemma 3.2 which generalizes inequality (1.23).

2 Necessary and sufficient conditions for norm-entropy inequalities.

Let us begin with a remark on Assumption (A_2).

Lemma 2.1. *Suppose that Φ contains a function φ_0 which is not μ-almost everywhere constant. If μ satisfies the inequality*

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \alpha\left(\|\nu-\mu\|_\Phi^*\right) \leq \mathrm{H}(\nu\mid\mu),$$

then α satisfies Assumption (A_2).

Proof. (See also [12], Proposition 2) Let us define $\Lambda_{\varphi_0}(s)=\log\int_{\mathcal{X}} e^{s\varphi_0}\,d\mu$, for all $s\in\mathbb{R}$. According to Theorem 1.10, we have $\forall s\geq 0,\quad \Lambda_{\varphi_0}(s)-s\langle\varphi_0,\mu\rangle \leq \alpha^{\circledast}(s)$. It is well known that $\lim_{s\to 0^+}\frac{\Lambda_{\varphi_0}(s)-s\langle\varphi_0,\mu\rangle}{s^2}=\frac{1}{2}\operatorname{Var}_\mu(\varphi_0)>0$. From this follows that $\liminf_{s\to 0^+}\dfrac{\alpha^{\circledast}(s)}{s^2}>0$, which easily implies (1.6). $\square$

Remark 2.2. *Note that if all the elements of Φ are μ-almost everywhere constant, then $\|\nu-\mu\|_\Phi^*=0$ for all $\nu\ll\mu$. Inequality (1.1) is thus satisfied, for all $\alpha\in\mathcal{C}$.*

The rest of this section is devoted to the proof of Theorem 1.7. The following lemma will be useful in the sequel :

Lemma 2.3. *Let X be a random variable such that $\mathbb{E}\left[e^{\delta|X|}\right]<+\infty$, for some $\delta>0$. Let us denote by Λ_X the Log-Laplace of X, which is defined by $\Lambda_X(s)=\log\mathbb{E}\left[e^{sX}\right]$, and by Λ_X^* its Cramér transform defined by $\Lambda_X^*(t)=\sup_{s\in\mathbb{R}}\{st-\Lambda_X(s)\}$, then the following upper-bound holds :*

$$\forall \varepsilon\in[0,1[,\quad \mathbb{E}\left[e^{\varepsilon\Lambda_X^*(X)}\right]\leq\frac{1+\varepsilon}{1-\varepsilon}.$$

Proof. (See also Lemma 5.1.14 of [7].) Let $a < b$ with $a \in \mathbb{R}\cup\{-\infty\}$ and $b \in \mathbb{R}\cup\{+\infty\}$ be the endpoints of dom Λ_X^*. Since Λ_X^* is convex and lower semi-continuous, $\{\Lambda_X^* \le t\}$ is an interval with endpoints $a \le a(t) \le b(t) \le b$, for all $t \ge 0$. As a consequence,

$$\forall t \ge 0, \quad \mathbb{P}(\Lambda_X^*(X) > t) = \mathbb{P}(X < a(t)) + \mathbb{P}(X > b(t)).$$

Let $m = \mathbb{E}[X]$. Since $\Lambda_X^*(m) = 0$, $a(t) \le m$. But for all $u \le m$, it is well known that

$$\mathbb{P}(X \le u) \le \exp(-\Lambda_X^*(u)) \tag{2.4}$$

If $a(t) > a$, the continuity of Λ_X^* on $]a, b[$ easily implies that $\Lambda_X^*(a(t)) = t$. Thus, according to (2.4),

$$\mathbb{P}(X < a(t)) \le e^{-t}.$$

If $a(t) = a$, then

$$\mathbb{P}(X < a) = \lim_{n\to+\infty} \mathbb{P}(X < a - 1/n) \overset{(i)}{\le} \lim_{n\to+\infty} \exp(-\Lambda_X^*(a - 1/n)) \overset{(ii)}{=} \lim_{n\to+\infty} 0 = 0,$$

where (i) comes from (2.4) and (ii) from $a - 1/n \notin$ dom Λ_X^*.
Therefore, in all cases $\mathbb{P}(X < a(t)) \le e^{-t}$. In the same way, we have $\mathbb{P}(X > b(t)) \le e^{-t}$. As a consequence,

$$\forall t \ge 0, \quad \mathbb{P}\left(\Lambda_X^*(X) > t\right) \le 2e^{-t}. \tag{2.5}$$

Finally, integrating by parts and using (2.5) in $(*)$ bellow, we get

$$\begin{aligned}\mathbb{E}\left[e^{\varepsilon\Lambda_X^*(X)}\right] &= \int_{-\infty}^{+\infty} e^t\mathbb{P}\left(\Lambda_X^*(X) > t/\varepsilon\right) dt = \int_{-\infty}^{0} e^t\, dt + \int_0^{+\infty} e^t\mathbb{P}(\Lambda_X^*(X) > t/\varepsilon)\, dt \\ &\overset{(*)}{\le} 1 + 2\int_0^{+\infty} e^{(1-1/\varepsilon)t}\, dt = \frac{1+\varepsilon}{1-\varepsilon}.\end{aligned}$$

□

Now, let us prove Theorem 1.7.

Proof of Theorem 1.7. Let us show that (1) implies (2). For $\varphi \in \Phi$, according to Theorem 1.10 and using the fact that $-\varphi \in \Phi$, we have

$$\forall s \in \mathbb{R}, \quad \log \int_{\mathcal{X}} e^{s(\varphi - \langle\varphi,\mu\rangle}\, d\mu \le \alpha^{\circledast}(|as|). \tag{2.6}$$

Define $\widetilde{\varphi} := \varphi - \langle\varphi,\mu\rangle$ and $\Lambda_{\widetilde{\varphi}}(s) := \log\int_{\mathcal{X}} e^{s(\varphi-\langle\varphi,\mu\rangle}\, d\mu$. Equation (2.6) immediately yields

$$\forall t \in \mathbb{R}, \quad \alpha\left(\frac{|t|}{a}\right) = \sup_{s\in\mathbb{R}}\left\{st - \alpha^{\circledast}(|as|)\right\} \le \sup_{s\in\mathbb{R}}\left\{st - \Lambda_{\widetilde{\varphi}}(s)\right\} = \Lambda_{\widetilde{\varphi}}^*(t).$$

According to Lemma 2.3, $\int_{\mathcal{X}} e^{\varepsilon\Lambda_{\widetilde{\varphi}}^*(\widetilde{\varphi})}\, d\mu \le \frac{1+\varepsilon}{1-\varepsilon}$, for all $\varepsilon \in [0, 1[$. Thus $\int_{\mathcal{X}} e^{\varepsilon\alpha\left(\frac{\widetilde{\varphi}}{a}\right)}\, d\mu \le \frac{1+\varepsilon}{1-\varepsilon}$. Since $\alpha\left(\frac{|.|}{a}\right)$ is convex and $\alpha(0) = 0$, we have $\alpha\left(\frac{\varepsilon|t|}{a}\right) \le \varepsilon\alpha\left(\frac{|t|}{a}\right)$. Therefore, $\int_{\mathcal{X}} e^{\alpha\left(\frac{\varepsilon|\widetilde{\varphi}|}{a}\right)} d\mu \le \frac{1+\varepsilon}{1-\varepsilon}$. In other words,

$$\forall\varphi \in \Phi, \quad \forall\varepsilon \in [0, 1[, \quad \int_{\mathcal{X}} \tau_\alpha\left(\frac{\varepsilon\widetilde{\varphi}}{a}\right) d\mu \le \frac{2\varepsilon}{1-\varepsilon}.$$

It is now easy to see that $\|\widetilde{\varphi}\|_{\tau_\alpha} \leq 3a$, for all $\varphi \in \Phi$.

Now let us show that (2) implies (1). According to Theorem 1.12,

$$\forall s \geq 0, \quad \int_{\mathcal{X}} e^{s\varphi}\, d\mu \leq e^{s\langle \varphi,\mu\rangle + \alpha^{\circledast}\left(\sqrt{2}m_\alpha \|\varphi - \langle\varphi,\mu\rangle\|_{\tau_\alpha} s\right)},$$

for all $\varphi \in \Phi$. As it is assumed that $\|\varphi - \langle \varphi, \mu\rangle\|_{\tau_\alpha} \leq M$, for all $\varphi \in \Phi$, we thus have

$$\forall \varphi \in \Phi, \quad \forall s \geq 0, \quad \int_{\mathcal{X}} e^{s\varphi}\, d\mu \leq e^{s\langle\varphi,\mu\rangle + \alpha^{\circledast}(as)},$$

with $a = \sqrt{2}m_\alpha M$. According to Theorem 1.10, this implies that μ satisfies the inequality

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \alpha\left(\frac{\|\nu - \mu\|_\Phi^*}{a}\right) \leq \mathrm{H}(\nu \mid \mu).$$

□

Example : Weighted Pinsker inequalities. Let $\chi : \mathcal{X} \to \mathbb{R}^+$ be a measurable function and let Φ_χ be the set of bounded measurable functions φ on $\mathcal{X}$ such that $|\varphi| \leq \chi$. In this framework, it is easily seen that

$$\|\nu - \mu\|_{\Phi_\chi}^* = \|\chi \cdot (\nu - \mu)\|_{TV},$$

where $\|\gamma\|_{TV}$ denotes the total-variation of the signed measure γ.

Theorem 2.7. *Suppose that $\int_{\mathcal{X}} \chi\, d\mu < +\infty$ and that $\alpha \in \mathcal{C}$ satisfies Assumptions (A_1) and (A_2), then the following propositions are equivalent :*

1. $\exists a > 0$, *such that* $\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \alpha\left(\dfrac{\|\chi\cdot(\nu - \mu)\|_{TV}}{a}\right) \leq \mathrm{H}(\nu \mid \mu),$
2. $\chi \in \mathbb{L}_{\tau_\alpha}(\mathcal{X}, \mu)$.

More precisely, if $\chi \in \mathbb{L}_{\tau_\alpha}(\mathcal{X},\mu)$, then one can take $a = 2\sqrt{2}m_\alpha\|\chi\|_{\tau_\alpha}$. Conversely, if (1) holds true, then

$$\|\chi\|_{\tau_\alpha} \leq \begin{cases} 3a, & \textit{if } \mu \textit{ has no atoms} \\ 3a + \int_{\mathcal{X}} \chi\, d\mu \cdot \|\mathbb{1}\|_{\tau_\alpha}, & \textit{otherwise} \end{cases}$$

Furthermore, the Luxemburg norm $\|\chi\|_{\tau_\alpha}$ can be estimated in the following way :

- *If* $\operatorname{dom}\alpha = \mathbb{R}^+$, *then* $\|\chi\|_{\tau_\alpha} \leq \inf_{\delta > 0}\left\{\dfrac{1}{\delta}\left(1 + \dfrac{\log\int_{\mathcal{X}} e^{\alpha(\delta\chi)}d\mu}{\log 2}\right)\right\}$
- *If* $\operatorname{dom}\alpha = [0, r_\alpha[$ *or* $[0, r_\alpha]$, *then* $\mathbb{L}_{\tau_\alpha}(\mathcal{X},\mu) = \mathbb{L}_\infty(\mathcal{X},\mu)$ *and*

$$r_\alpha^{-1}\|\chi\|_\infty \leq \|\chi\|_{\tau_\alpha} \leq \sup\{t > 0 : \alpha(t) \leq \log 2\}^{-1} \cdot \|\chi\|_\infty.$$

Remark 2.8. *If $\alpha \in \mathcal{C}$ satisfies Assumptions (A_1) and (A_2) and is such that* $\operatorname{dom} \alpha = \mathbb{R}^+$, *we have thus shown the following weighted version of Pinsker inequality :*

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \|\chi \cdot (\nu - \mu)\|_{TV} \leq 2\sqrt{2} m_\alpha \inf_{\delta > 0} \left\{ \frac{1}{\delta} \left(1 + \frac{\log \int_{\mathcal{X}} e^{\alpha(\delta \chi)} d\mu}{\log 2} \right) \right\} \alpha^{-1} \left(\mathrm{H}(\nu \mid \mu) \right) \quad (2.9)$$

Inequality (2.9) completely extends Bolley and Villani's results (1.21) and (1.22). The proof of Bolley and Villani is very different from ours. Roughly speaking, it relies on a direct comparison of the two integrals $\int_{\mathcal{X}} \chi \left| \frac{d\nu}{d\mu} - 1 \right| d\mu$ and $\int_{\mathcal{X}} \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} d\mu$.

Proof of Theorem 2.7. According to Theorem 1.7, it suffices to show that

$$2\|\chi\|_{\tau_\alpha} \geq \sup_{\varphi \in \Phi_\chi} \{ \|\varphi - \langle \varphi, \mu \rangle \|_{\tau_\alpha} \} \geq \begin{cases} \|\chi\|_{\tau_\alpha} & \text{if } \mu \text{ is non-atomic} \\ \|\chi\|_{\tau_\alpha} - \int_{\mathcal{X}} \chi \, d\mu \cdot \|\mathbb{1}\|_{\tau_\alpha} & \text{otherwise.} \end{cases} \quad (2.10)$$

Let us prove the first inequality of (2.10) : If $\varphi \in \Phi_\chi$, then $|\varphi| \leq \chi$, thus $\|\varphi - \langle \varphi, \mu \rangle\|_{\tau_\alpha} \leq \|\chi\|_{\tau_\alpha} + \|\langle \varphi, \mu \rangle\|_{\tau_\alpha}$. Thanks to Jensen inequality, for all $\lambda > 0$, we have $\int_{\mathcal{X}} \tau_\alpha \left(\frac{\langle \varphi, \mu \rangle}{\lambda} \right) d\mu \leq \int_{\mathcal{X}} \tau_\alpha \left(\frac{\varphi}{\lambda} \right) d\mu$. Thus, $\|\langle \varphi, \mu \rangle\|_{\tau_\alpha} \leq \|\varphi\|_{\tau_\alpha}$, which proves the desired inequality.

Thanks to triangle inequality $\sup_{\varphi \in \Phi_\chi} \|\varphi - \langle \varphi, \mu \rangle\|_{\tau_\alpha} \geq \|\chi - \langle \chi, \mu \rangle\|_{\tau_\alpha} \geq \|\chi\|_{\tau_\alpha} - \| \int_{\mathcal{X}} \chi \, d\mu \|_{\tau_\alpha} = \|\chi\|_{\tau_\alpha} - \int_{\mathcal{X}} \chi \, d\mu \cdot \|\mathbb{1}\|_{\tau_\alpha}$.

Suppose that μ has no atoms, then $\chi \cdot \mu$ has no atoms too. As a consequence, there exists a measurable set $A \subset \mathcal{X}$ such that $\int_A \chi \, d\mu = \frac{1}{2} \int_{\mathcal{X}} \chi \, d\mu$. Define $\widetilde{\chi} = \chi \mathbb{1}_A - \chi \mathbb{1}_{A^c}$. Then $|\widetilde{\chi}| = \chi$ and $\langle \widetilde{\chi}, \mu \rangle = 0$. Thus $\sup_{\varphi \in \Phi_\chi} \|\varphi - \langle \varphi, \mu \rangle\|_{\tau_\alpha} \geq \|\widetilde{\chi} - \langle \widetilde{\chi}, \mu \rangle\|_{\tau_\alpha} = \|\widetilde{\chi}\|_{\tau_\alpha} = \|\chi\|_{\tau_\alpha}$.

Now, let us explain how to majorize the Luxemburg norms. Suppose that $\operatorname{dom} \alpha = \mathbb{R}^+$.

Let $\delta > 0$ be fixed and assume that $\|\chi\|_{\tau_\alpha} \geq \frac{1}{\delta}$ and that $\displaystyle\int_{\mathcal{X}} e^{\alpha(\delta \chi)} \, d\mu < +\infty$. Then, denoting $\lambda = \|\chi\|_{\tau_\alpha}$, we have

$$2^{\delta \lambda} \overset{(i)}{=} \left[\int_{\mathcal{X}} \exp \alpha \left(\frac{\chi}{\lambda} \right) d\mu \right]^{\delta \lambda} \overset{(ii)}{\leq} \int_{\mathcal{X}} \exp \delta \lambda \alpha \left(\frac{\chi}{\lambda} \right) d\mu \overset{(iii)}{\leq} \int_{\mathcal{X}} \exp \alpha \left(\delta \chi \right) d\mu$$

where (i) come from the definition of $\lambda = \|\chi\|_{\tau_\alpha}$, (ii) from Jensen inequality and (iii) from the inequality $\alpha(x/M) \leq \alpha(x)/M$, for all $M \geq 1$. Taking the log in both side of the above inequality yields $\lambda \leq \frac{1}{\delta \log 2} \int_{\mathcal{X}} \exp \alpha \left(\delta \chi \right) d\mu$ and *a fortiori*,

$$\lambda \leq \frac{1}{\delta} + \frac{1}{\delta \log 2} \int_{\mathcal{X}} \exp \alpha \left(\delta \chi \right) d\mu.$$

If $\|\chi\|_{\tau_\alpha} \leq \frac{1}{\delta}$ or if $\displaystyle\int_{\mathcal{X}} e^{\alpha(\delta \chi)} \, d\mu = +\infty$, the preceding inequality remains true. Optimizing in $\delta > 0$ gives the desired result.

The case where $\operatorname{dom} \alpha$ is a bounded interval is left to the reader. □

Remark 2.11. *It is easy to show that when $\alpha(x) = x^2$, the Luxemburg norm $\|\chi\|_{\tau_{x^2}}$ can be estimated in the following way :*

$$\|\chi\|_{\tau_{x^2}} \leq \inf_{\delta > 0} \frac{1}{\delta} \sqrt{1 + \frac{\log \int_{\mathcal{X}} e^{\delta^2 \chi^2} \, d\mu}{\log 2}}.$$

With this upper-bound, and using the fact that $m_{x^2} = 2e$ (left to the reader), one obtains

$$\|\chi\cdot(\nu-\mu)\|_{TV} \le 4e \inf_{\delta>0} \frac{1}{\delta}\sqrt{1+\frac{\log\int_{\mathcal{X}} e^{\delta^2\chi^2}\,d\mu}{\log 2}}\cdot\sqrt{2\,\mathrm{H}(\nu\mid\mu)}, \tag{2.12}$$

which differs from (1.22) only by numerical factors. However the proof of (2.12) relies on Theorem 1.12 which is a non trivial result. In the following proposition, we improve the constants in (2.12), using this time only elementary computations.

Proposition 2.13. *For every measurable function $\chi : \mathcal{X} \to \mathbb{R}^+$, the following inequality holds*

$$\|\chi\cdot(\nu-\mu)\|_{TV} \le \inf_{\delta>0} \frac{1}{\delta}\sqrt{1+4\log\int_{\mathcal{X}} e^{\delta^2\chi^2}\,d\mu}\cdot\sqrt{2\,\mathrm{H}(\nu\mid\mu)}. \tag{2.14}$$

Proof. First let us show that if X is a real random variable such that $\mathbb{E}\left[e^{X^2}\right] < +\infty$ one has the following upper bound :

$$\forall s\ge 0,\quad \mathbb{E}\left[e^{s(X-\mathbb{E}[X])}\right] \le e^{s^2/2}\cdot\mathbb{E}\left[e^{X^2}\right]^{2s^2}. \tag{2.15}$$

Let $\widetilde{X}$ be an independent copy of X. According to Jensen inequality, we have $\mathbb{E}\left[e^{s(X-\mathbb{E}[X])}\right] \le \mathbb{E}\left[e^{s(X-\widetilde{X})}\right]$. The random variable $X-\widetilde{X}$ is symmetric, thus $\mathbb{E}\left[(X-\widetilde{X})^{2k+1}\right] = 0$, for all k. Consequently,

$$\begin{aligned}\mathbb{E}\left[e^{s(X-\mathbb{E}[X])}\right] \le \mathbb{E}\left[e^{s(X-\widetilde{X})}\right] &= \sum_{k=0}^{+\infty}\frac{s^{2k}\mathbb{E}\left[(X-\widetilde{X})^{2k}\right]}{(2k)!}\\ &\le \sum_{k=0}^{+\infty}\frac{s^{2k}\mathbb{E}\left[(X-\widetilde{X})^{2k}\right]}{2^k\cdot k!} = \mathbb{E}\left[e^{s^2(X-\widetilde{X})^2/2}\right].\end{aligned}$$

It is easily seen that $\mathbb{E}\left[e^{s^2(X-\widetilde{X})^2/2}\right] \le \mathbb{E}\left[e^{s^2X^2}\right]^2$, and if $s\le 1$, $\mathbb{E}\left[e^{s^2X^2}\right]^2 \le \mathbb{E}\left[e^{X^2}\right]^{2s^2}$. Hence,

$$\forall s\le 1,\quad \mathbb{E}\left[e^{s(X-\mathbb{E}[X])}\right] \le \mathbb{E}\left[e^{X^2}\right]^{2s^2}.$$

But if $s\ge 1$, one has

$$\mathbb{E}\left[e^{s(X-\mathbb{E}[X])}\right] \le \mathbb{E}\left[e^{s(X-\widetilde{X})}\right] \le \mathbb{E}\left[e^{s^2/2+(X-\widetilde{X})^2/2}\right] \le e^{s^2/2}\cdot\mathbb{E}\left[e^{X^2}\right]^2 \le e^{s^2/2}\cdot\mathbb{E}\left[e^{X^2}\right]^{2s^2}.$$

So, the inequality $\mathbb{E}\left[e^{s(X-\mathbb{E}[X])}\right] \le e^{s^2/2}\cdot\mathbb{E}\left[e^{X^2}\right]^{2s^2}$ holds for all $s\ge 0$.

Let φ be a bounded measurable function such that $|\varphi|\le\chi$. Applying inequality (2.15), one obtains immediately

$$\int_{\mathcal{X}} e^{s(\varphi-\langle\varphi,\mu\rangle)}\,d\mu \le e^{s^2M^2/2},$$

with $M = \sqrt{1 + 4\log \int_{\mathcal{X}} e^{\chi^2}\, d\mu}$. Thus, according to Theorem 1.10 the following norm-entropy inequality holds :

$$\|\chi \cdot (\nu - \mu)\|_{TV} \leq \sqrt{1 + 4\log \int_{\mathcal{X}} e^{\chi^2}\, d\mu} \cdot \sqrt{2\,\mathrm{H}(\nu \mid \mu)}.$$

Replacing χ by $\delta\chi$ and using homogeneity one obtains (2.14). □

Remark 2.16. *Note that (2.14) is sharper than (2.12). But (1.22) is still sharper than (2.14).*

3 Applications to transportation cost inequalities.

In this section, we will see how to derive transportation-cost inequalities from norm-entropy inequalities. Let us begin with the proof of Theorem 1.13.

Proof of Theorem 1.13. First let us show that (1) implies (2). According to Theorem 1.7, one has $\sup_{\varphi \in \mathrm{BLip}_1(\mathcal{X},d)} \|\varphi - \langle \varphi, \mu\rangle\|_{\tau_\alpha} \leq 3a$. In particular, using an easy approximation technique, $\|d(x_0,\,.\,) - \langle d(x_0,\,.\,), \mu\rangle\|_{\tau_\alpha} \leq 3a$, and thus $d(x_0,\,.\,) \in \mathbb{L}_{\tau_\alpha}(\mathcal{X}, \mu)$.

Now let us see that (2) implies (1). Let $x_0 \in \mathcal{X}$; observe that $\mathcal{T}_d(\nu, \mu) = \|\nu - \mu\|_{\Phi_{x_0}}$, with $\Phi_{x_0} = \{\varphi \in \mathrm{BLip}_1(\mathcal{X}, d) : \varphi(x_0) = 0\}$. But $\Phi_{x_0} \subset \widetilde{\Phi}_{x_0} := \{\varphi : \forall x \in \mathcal{X}, |\varphi(x)| \leq d(x_0, x)\}$. Thus, $\mathcal{T}_d(\nu, \mu) \leq \|\nu - \mu\|_{\widetilde{\Phi}_{x_0}} = \|d(x_0,\,.\,) \cdot (\nu - \mu)\|_{TV}$. Applying Theorem 2.7, one concludes that if $d(x_0,\,.\,) \in \mathbb{L}_{\tau_\alpha}(\mathcal{X}, \mu)$, then the inequality $\forall \nu \in \mathcal{P}(\mathcal{X}),\quad \alpha\left(\frac{\mathcal{T}_d(\nu,\mu)}{a}\right) \leq \mathrm{H}(\nu \mid \mu)$ holds with $a = 2\sqrt{2}m_\alpha \|d(x_0,\,.\,)\|_{\tau_\alpha}$. As this is true for all $x_0 \in \mathcal{X}$, the same inequality holds for $a = 2\sqrt{2}m_\alpha \inf_{x_0 \in \mathcal{X}} \|d(x_0,\,.\,)\|_{\tau_\alpha}$. □

When the cost function is of the form $c(x, y) = q(d(x, y))$, we will use the following result which is adapted from Proposition 7.10 of [18] :

Lemma 3.1. *Let c be a cost function on $\mathcal{X}$ of the form $c(x, y) = q(d(x, y))$, with $q : \mathbb{R}^+ \to \mathbb{R}^+$ an increasing convex function. Let $x_0 \in \mathcal{X}$ and define $\chi_{x_0}(x) = \frac{1}{2}q(2d(x, x_0))$, for all $x \in \mathcal{X}$. Then the following inequality holds :*

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad q\left(\mathcal{T}_d(\nu, \mu)\right) \leq \mathcal{T}_c(\nu, \mu) \leq \|\chi_{x_0} \cdot (\nu - \mu)\|_{TV}. \tag{3.2}$$

Proof. Applying Jensen inequality, one gets $q\left(\int_{\mathcal{X}^2} d(x, y)\, d\pi(x, y)\right) \leq \int_{\mathcal{X}^2} q(d(x, y))\, d\pi(x, y)$, for all $\pi \in \Pi(\nu, \mu)$. Thus according to the definition of $\mathcal{T}_c(\nu, \mu)$ (see (1.3)), one deduces immediately the first inequality in (3.2). It follows from the triangle inequality and the convexity of q that

$$c(x, y) = q(d(x, y)) \leq q(d(x, x_0) + d(y, y_0)) \leq \frac{1}{2}\left[q(2d(x, x_0)) + q(2d(y, x_0))\right] = \chi_{x_0}(x) + \chi_{x_0}(y).$$

Thus $c(x, y) \leq d_{\chi_{x_0}}(x, y)$, with $d_{\chi_{x_0}}(x, y) = (\chi_{x_0}(x) + \chi_{x_0}(y))\mathbb{1}_{\{x \neq y\}}$ and consequently $\mathcal{T}_c(\nu, \mu) \leq \mathcal{T}_{d_{\chi_{x_0}}}(\nu, \mu)$. But $\mathcal{T}_{d_{\chi_{x_0}}}(\nu, \mu) = \|\chi_{x_0} \cdot (\nu - \mu)\|_{TV}$ (see for instance, Prop. VI.7 p. 154 of [11]), which proves the second part of (3.2). □

Using the second part of inequality (3.2) together with Theorem 2.7, one immediately derives the following result which is the first half of Theorem 1.15 :

Proposition 3.3. *Let c be a cost function on $\mathcal{X}$ of the form $c(x,y) = q(d(x,y))$, with $q : \mathbb{R}^+ \to \mathbb{R}^+$ an increasing convex function and $\alpha \in \mathcal{C}$ satisfying Assumptions (A_1) and (A_2). Then the following T.C.I holds*

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \alpha\left(\frac{\mathcal{T}_c(\nu,\mu)}{a}\right) \leq \mathrm{H}(\nu \mid \mu), \tag{3.4}$$

with $a = \sqrt{2}m_\alpha \inf_{x_0\in\mathcal{X}} \|q(2d(x_0,\,.\,))\|_{\tau_\alpha}$. Furthermore, if q satisfies the Δ_2-condition (1.16) with constant $K > 0$, then one can take $a = \sqrt{2}Km_\alpha \inf_{x_0\in\mathcal{X}} \|c(x_0,\,.\,)\|_{\tau_\alpha}$.

Remark 3.5. *If q satisfies the Δ_2-condition and* dom $\alpha = \mathbb{R}^+$, *then μ satisfies the following T.C.I :*

$$\forall \nu \in \mathcal{P}(\mathcal{X}), \quad \mathcal{T}_c(\nu,\mu) \leq \sqrt{2}Km_\alpha \inf_{x_0\in\mathcal{X},\,\delta>0} \frac{1}{\delta}\left(1 + \frac{\log \int_{\mathcal{X}} e^{\delta\alpha(c(x_0,x))}\, d\mu(x)}{\log 2}\right) \alpha^{-1}\left(\mathrm{H}(\nu \mid \mu)\right)$$

Now, let us prove the second half of Theorem 1.15 :

Proposition 3.6. *Let c be a cost function on $\mathcal{X}$ of the form $c(x,y) = q(d(x,y))$, with $q : \mathbb{R}^+ \to \mathbb{R}^+$ an increasing convex function satisfying the Δ_2-condition (1.16) with a constant $K > 0$ and let $\alpha \in \mathcal{C}$ satisfy Assumption (A_1). If $\int_{\mathcal{X}} c(x_0,x)\,d\mu(x) < +\infty$ for all $x_0 \in \mathcal{X}$ and if the T.C.I (3.4) holds for some $a > 0$, then the function $c(x_0,\,.\,)$ belongs to $\mathbb{L}_{\tau_\alpha}(\mathcal{X},\mu)$ for all $x_0 \in \mathcal{X}$.*

Proof. According to the first part of inequality (3.2), $q\left(\mathcal{T}_d(\nu,\mu)\right) \leq \mathcal{T}_c(\nu,\mu)$, thus, if (3.4) holds for some $a > 0$, then $\widetilde{\alpha}\left(\mathcal{T}_d(\nu,\mu)\right) \leq \mathrm{H}(\nu \mid \mu)$, for all $\nu \in \mathcal{P}(\mathcal{X})$, where $\widetilde{\alpha}(x) = \alpha\left(\frac{q(x)}{a}\right)$. According to Theorem 1.7, this implies that $\sup_{\varphi\in\mathrm{BLip}_1(\mathcal{X},d)} \|\varphi - \langle\varphi,\mu\rangle\|_{\tau_{\widetilde{\alpha}}} \leq 3$. In particular, using an easy approximation argument, it is easy to see that $\|d(x_0,\,.\,) - \langle d(x_0,\,.\,),\mu\rangle\|_{\tau_{\widetilde{\alpha}}} \leq 3$, which implies that $d(x_0,\,.\,) \in \mathbb{L}_{\tau_{\widetilde{\alpha}}}(\mathcal{X},\mu)$. Let $\lambda > 0$ be such that $\int_{\mathcal{X}} \tau_{\widetilde{\alpha}}\left(\frac{d(x_0,x)}{\lambda}\right) d\mu(x) < +\infty$ and let n be a positive integer such that $2^n \geq \lambda$. Then, according to the Δ_2 condition satisfied by q, one has $q\left(\frac{x}{\lambda}\right) \geq q\left(\frac{x}{2^n}\right) \geq \frac{1}{K^n}q(x)$, for all $x \in \mathbb{R}^+$. Consequently, $\tau_{\widetilde{\alpha}}\left(\frac{x}{\lambda}\right) \geq \tau_\alpha\left(\frac{q(x)}{aK^n}\right)$, for all $x \in \mathbb{R}^+$. From this follows that

$$\int_{\mathcal{X}} \tau_\alpha\left(\frac{c(x_0,x)}{aK^n}\right) d\mu(x) = \int_{\mathcal{X}} \tau_\alpha\left(\frac{q(d(x,x_0))}{aK^n}\right) d\mu(x) \leq \int_{\mathcal{X}} \tau_{\widetilde{\alpha}}\left(\frac{d(x_0,x)}{\lambda}\right) d\mu(x) < +\infty$$

and thus $c(x_0,\,.\,) \in \mathbb{L}_{\tau_\alpha}(\mathcal{X},\mu)$. ☐

Proof of Theorem 1.15. Theorem 1.15 follows immediately from Propositions 3.3 and 3.6. ☐

References

[1] S. G. Bobkov, I. Gentil, and M. Ledoux. Hypercontractivity of Hamilton-Jacobi equations. *Journal de Mathématiques Pures et Appliquées*, 80(7):669–696, 2001.

[2] F. Bolley, A. Guillin, and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. preprint., 2005.

[3] F. Bolley and C. Villani. Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. *Annales de la Faculté des Sciences de Toulouse.*, 14:331–352, 2005.

[4] V. V. Buldygin and Yu.V. Kozachenko. *Metric characterization of random variables and random processes.* American Mathematical Society, 2000.

[5] P. Cattiaux and N. Gozlan. Deviations bounds and conditional principles for thin sets. preprint., 2005.

[6] P. Cattiaux and A. Guillin. Talagrand's like quadratic transportation cost inequalities. preprint., 2004.

[7] A. Dembo and O. Zeitouni. *Large deviations techniques and applications. Second edition.* Applications of Mathematics 38. Springer Verlag, 1998.

[8] H. Djellout, A. Guillin, and L. Wu. Transportation cost-information inequalities for random dynamical systems and diffusions. *Annals of Probability*, 32(3B):2702–2732, 2004.

[9] I. Gentil, A. Guillin, and L. Miclo. Modified logarithmic sobolev inequalities and transportation inequalities. preprint., 2005.

[10] N. Gozlan. Conditional principles for random weighted measures. *ESAIM P&S*, 9:283–306, 2005.

[11] N. Gozlan. Principe conditionnel de Gibbs pour des contraintes fines approchées et inégalités de transport. PhD Thesis, Université de Paris 10., 2005.

[12] N. Gozlan and C. Léonard. A large deviation approach to some transportation cost inequalities. preprint., 2005.

[13] Yu.V. Kozachenko and E.I. Ostrovskii. Banach spaces of random variables of sub-gaussian type. *Theor. Probability and Math. Statist.*, 3.:45–56, 1986.

[14] K. Marton. A simple proof of the blowing-up lemma. *IEEE Transactions on Information Theory*, 32:445–446, 1986.

[15] K. Marton. Bounding $\bar{d}$-distance by informational divergence: a way to prove measure concentration. *Annals of Probability*, 24:857–866, 1996.

[16] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *Journal of Functional Analysis*, 173:361–400, 2000.

[17] M. Talagrand. Transportation cost for gaussian and other product measures. *Geometric and Functional Analysis*, 6:587–600, 1996.

[18] C. Villani. *Topics in Optimal Transportation.* Graduate Studies in Mathematics 58. American Mathematical Society, Providence RI, 2003.