A LARGE DEVIATION APPROACH TO SOME TRANSPORTATION COST INEQUALITIES

NATHAEL GOZLAN AND CHRISTIAN LÉONARD

ABSTRACT. New transportation cost inequalities are derived by means of elementary large deviation reasonings. Their dual characterization is proved; this provides an extension of a well-known result of S. Bobkov and F. Götze. Their tensorization properties are investigated. Sufficient conditions (and necessary conditions too) for these inequalities are stated in terms of the integrability of the reference measure. Applying these results leads to new deviation results: concentration of measure and deviations of empirical processes.

1. INTRODUCTION

In the whole paper, $\mathcal{X}$ is a Polish space equipped with its Borel σ-field. We denote $\mathcal{P}(\mathcal{X})$ the set of all probability measures on $\mathcal{X}$.

1.1. Transportation cost inequalities and concentration of measure.

Let us first recall what transportation cost inequalites are and their well known consequences in terms of concentration of measure.

Transportation cost. Let $c : \mathcal{X} \times \mathcal{X} \to [0, \infty)$ be a measurable function on the product space $\mathcal{X} \times \mathcal{X}$. For any couple of probability measures μ and ν on $\mathcal{X}$, the transportation cost (associated with the cost function c) of μ on ν is

$$\mathcal{T}_c(\mu, \nu) = \inf_\pi \int_{\mathcal{X}\times\mathcal{X}} c(x, y)\, \pi(dxdy) \in [0, \infty]$$

where the inf is taken over all probability measures π on $\mathcal{X} \times \mathcal{X}$ with first marginal $\pi(dx \times \mathcal{X}) = \mu(dx)$ and second marginal $\pi(\mathcal{X} \times dy) = \nu(dy)$.

T_p-inequalities. Popular cost functions are $c(x, y) = d(x, y)^p$ where d is a metric on $\mathcal{X}$ and $p \geq 1$. It is known that for some $\mu \in \mathcal{P}(\mathcal{X})$ and $p \geq 1$ one can prove the following *transportation cost inequality*

$$\mathcal{T}_{d^p}(\mu, \nu)^{1/p} \leq \sqrt{2CH(\nu \mid \mu)}, \quad \forall \nu \in \mathcal{P}(\mathcal{X}) \tag{1.1}$$

for some positive constant C, where $H(\nu \mid \mu)$ is the *relative entropy* of ν with respect to μ defined by

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

if ν is absolutely continuous with respect to μ and $H(\nu \mid \mu) = \infty$ otherwise. In presence of the family of inequalities (1.1), one says that μ satifies $T_p(C)$.

Date: July 2005.

2000 *Mathematics Subject Classification.* 60E15, 60F10.

Key words and phrases. Transportation cost inequalities, Large deviations, Concentration of measure.

For instance, Csiszár-Kullback-Pinsker's inequality, see (2.9), is $T_1(1)$ with the Hamming's metric $d(x,y) = \mathbf{1}_{x\neq y.}$ Csiszár-Kullback-Pinsker's inequality is often called Pinsker's inequality, it will be refered later as CKP inequality. It holds for any $\mu \in \mathcal{P}(\mathcal{X})$. On the other hand, T_2-inequalities are much more difficult to obtain. It is shown in the articles by F. Otto and C. Villani [?] and by S. Bobkov, I. Gentil and M. Ledoux [1], that if μ satisfies the logarithmic Sobolev inequality, then it also satisfies T_2. A standard example of probability measure μ that satisfies T_2 is the normal law. In [16], M. Talagrand has given a proof of $T_2(C)$ for the standard normal law not relying on any log-Sobolev inequality, for the sharp constant $C = 1$.

Concentration of measure. As a consequence of $T_1(C)$, K. Marton [12, 13] has obtained the following *concentration inequality* for μ :

$$\mu(\{x; d(x,A) > r\}) \le \exp\left[-\left(\frac{r}{\sqrt{2C}} - \sqrt{\log 2}\right)^2\right] \tag{1.2}$$

for all measurable subset A such that $\mu(A) \ge 1/2$ and all $r \ge \sqrt{2C\log 2}$. *Marton's concentration argument* easily extends to more general situations. This is of considerable importance and justifies the search for T_1-inequalities.

Product of measures. Suppose that $\mu_1, \dots, \mu_n$ satisfy respectively $T_p(C_1), \dots, T_p(C_n)$. By means of a coupling argument which is also due to K. Marton [13] (the so-called *Marton's coupling argument*), one can check that when $p = 1$, the product measure $\mu_1 \otimes \cdots \otimes \mu_n$ satisfies $T_1(C_1 + \cdots + C_n)$, while when $p = 2$, $\mu_1 \otimes \cdots \otimes \mu_n$ satisfies $T_2(\max(C_1, \dots, C_n))$. In particular, if μ satisfies $T_1(C)$ then $\mu^{\otimes n}$ satisfies $T_1(nC)$. This inequality deteriorates as n grows. On the other hand, if μ satisfies $T_2(C)$ then $\mu^{\otimes n}$ also satisfies $T_2(C)$ and this still holds for the infinite product $\mu^{\otimes\infty}$.
By Jensen's inequality, we have $(\mathcal{T}_d)^2 \le \mathcal{T}_{d^2}$ so that $T_2(C)$ implies $T_1(C)$. As the standard normal law γ satisfies $T_2(1)$, it is also shown in [16] that the standard normal law on $\mathbb{R}^n$: γ^n, satisfies $T_2(1)$ and therefore $T_1(1)$ and the concentration inequality

$$\gamma^n(\{x; d(x,A) > r\}) \le \exp\left[-\left(\frac{r}{\sqrt{2}} - \sqrt{\log 2}\right)^2\right]$$

for all measurable subset A such that $\mu(A) \ge 1/2$ and all $r \ge \sqrt{2\log 2}$ where d is the Euclidean distance on $\mathbb{R}^n$. This concentration result holds for all n and is very close to the optimal concentration result obtained by means of isoperimetric arguments (see M. Ledoux's monograph [10], Corollary 2.6) which is: $\gamma^n(\{x; d(x,A) > r\}) \le e^{-r^2/2}$, for all $r \ge 0$.
In view of (1.2) and of this optimal concentration inequality, it now appears that with $\mathcal{X} = \mathbb{R}^n$, $T_1(C)$ implies that μ concentrates at least as a normal law with variance C. One may say that μ performs a Gaussian concentration when (1.2) holds for some C.

Criteria for T_1. It has recently been proved by H. Djellout, A. Guillin and L. Wu in [7] that μ satisfies $T_1(C)$ for some C if and only if

$$\int_{\mathcal{X}} e^{a_o d(x_o,x)^2}\,\mu(dx) < \infty \tag{1.3}$$

for some $a_o > 0$ and some (and therefore all) x_o in $\mathcal{X}$. It follows that *(1.3) is a characterization of the Gaussian concentration.* The proof of this result in [7] relies on a dual characterization of T_1 which has been obtained by S. Bobkov and F. Götze in [2]. This characterization is the following: $T_1(C)$ holds if and only if

$$\log \int_{\mathcal{X}} e^{s(\varphi - \langle \varphi, \mu \rangle)}\, d\mu \leq Cs^2/2, \tag{1.4}$$

for all $s \geq 0$ and all bounded Lipschitz function φ with $\|\varphi\|_{\mathrm{Lip}} \leq 1$.
The criterion (1.3) has been recovered very recently by F. Bolley and C. Villani in [**?**] where the relation between C and a_o is improved. This new proof relies on a strengthening of CKP inequality where weights are allowed in the total variation norm. For a statement of this strengthened CKP inequality, see Corollary 3.24 below.

1.2. **Presentation of the results.** In this article, a larger class of transportation cost inequalities is investigated. It appears that the transportation cost inequalities T_p defined by (1.1) enter the following larger class of inequalities, which will also be called transportation cost inequalities (TCIs):

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

where $\alpha : [0, \infty) \to [0, \infty)$ is an increasing[1] function which vanishes at 0. The inequality (1.1) corresponds $c = d^p$ with $\alpha(t) = t^{2/p}/(2C)$, $t \geq 0$. Of course, one should rigorously restrict (1.5) to those $\nu \in \mathcal{P}(\mathcal{X})$ such that $\mathcal{T}_c(\mu, \nu)$ is well-defined.
The aim of this paper is threefold.

(i) One proves TCIs by means of large deviation reasonings. The authors hope that this should provide a guideline for other functional inequalities.
(ii) One obtains deviation results by means of TCIs.
(iii) One extends already existing results, especially in the area of T_1-inequalities.

One says that we have a T_1-inequality if

$$\alpha(\mathcal{T}_d(\mu, \nu)) \leq H(\nu \mid \mu),\ \forall \nu \in \mathcal{P}_d(\mathcal{X}). \tag{T_1}$$

where d is a metric and $\mathcal{P}_d(\mathcal{X})$ is the set of all probability measures which integrate $d(x_o, x)$.

As regards item (i), it is no surprise that, because of the relative entropy entering TCIs, Sanov theorem plays a crucial role in our approach. Let

$$L_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$$

be the empirical measure of an n-iid sample (X_i) of the law $\mu \in \mathcal{P}(\mathcal{X})$. Sanov theorem states that the sequence $\{L_n\}_{n \geq 1}$ obeys the large deviation principle with rate function $\nu \mapsto H(\nu \mid \mu)$. The main idea is to control the deviations of the nonnegative random variables $\mathcal{T}_c(\mu, L_n)$ as n tends to infinity. An easy heuristic description of this program is displayed at Section 2.2. We obtain the

[1] In the whole paper, by an increasing function it is meant a nondecreasing function which may be constant on some intervals.

Recipe 1.6. Any increasing function α such that $\alpha(0) = 0$ and

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\mathcal{T}_c(\mu, L_n) \geq t) \leq -\alpha(t)$$

for all $t \geq 0$, satisfies the TCI (1.5).

Rigorously, one will have to require that α is a left continuous function. This result will be proved at Theorem 7.1 and a weak version of it (with α convex) is proved at Proposition 5.5.
Not only TCIs can be derived with this recipe but also another class of functional inequalities which we call Norm-Entropy Inequalities (NEIs), see (2.6) for their definition. Let us only emphasize in this introductory section that T_1-inequalities are NEIs.

As regards item (ii), **concentration inequalities** for general measures and **deviation inequalities for empirical processes** are derived by means of T_1-inequalities at Section 6.

As regards item (iii), the main technical (easy) result is Theorem 3.7 which is an extension of Bobkov and Götze's characterization of $T_1(C)$ stated at (1.4). It gives a **dual characterization** of all *convex* TCIs: those TCIs with α convex and increasing. Note that, up to the knowledge of the authors, all known TCIs are convex. As a consequence among others, one recovers the results of [**?**] about weighted CKP inequalities at Corollary 3.24.

Tensorization of convex TCIs is also handled. The main result on this topic is Theorem 4.2. It states that if $\alpha_1(\mathcal{T}_{c_1}(\mu_1, \nu_1)) \leq H(\nu_1 \mid \mu_1)$ for all ν_1 and $\alpha_2(\mathcal{T}_{c_2}(\mu_2, \nu_2)) \leq H(\nu_2 \mid \mu_2)$ for all ν_2, then $\alpha_1 \square \alpha_2(\mathcal{T}_{c_1 \oplus c_2}(\mu_1 \otimes \mu_2, \nu)) \leq H(\nu \mid \mu_1 \otimes \mu_2)$ for all ν probability measure on the product space, where $\alpha_1 \square \alpha_2$ is the inf-convolution of α_1 and α_2.

Integral criteria are investigated in Section 5. It emerges from our analysis via large deviations, that integral criteria only control the behavior of $\alpha(t)$ in (1.5) for t away from zero. As a consequence, complete results are only derived for T_1-inequalities. It is also proved that the function $\alpha(t)$ of a T_1-inequality has a quadratic behavior for t near zero. The integral criterion for T_1 is stated at Theorem 5.19. It is the following:
Let d be a lower semicontinuous metric. Suppose that $a \geq 0$ satisfies $\int_{\mathcal{X}} e^{ad(x_o,x)}\, \mu(dx) \leq 2$ for some $x_o \in \mathcal{X}$ and that γ is an increasing convex function which satisfies $\gamma(0) = 0$ and $\int_{\mathcal{X}} e^{\gamma(d(x_1,x))}\, \mu(dx) \leq B < \infty$ for some $x_1 \in \mathcal{X}$, then

$$\alpha(t) = \max\left((\sqrt{at+1} - 1)^2, 2\gamma(t/2) - 2\log B\right),\ t \geq 0$$

satisfies (T_1).
Note that $(\sqrt{at+1} - 1)^2 = a^2t^2/4 + o_{t\to 0}(t^2)$ is efficient for t near zero, while $2\gamma(t/2) - 2\log B$ is efficient for t away from zero.
This theorem extends the integral criterion (1.3) of [7] and [**?**].

The last Section 7 is devoted to abstract results. In particular, the extended version Recipe 2.8 of Recipe 1.6 is proved at Theorem 7.1. The authors hope that the set of abstract results stated in this section could be the starting point of the derivations of new functional inequalities.

Contents

2. Deriving $\mathcal{T}$-inequalities by means of large deviations. Heuristics

The dual equality associated with the primal minimization problem leading to $\mathcal{T}_c(\mu,\nu)$ is

$$\mathcal{T}_c(\mu,\nu) = \sup_{(\psi,\varphi)\in\Phi_c} \left\{ \int_{\mathcal{X}} \psi\, d\mu + \int_{\mathcal{X}} \varphi\, d\nu \right\} \tag{2.1}$$

where Φ_c is the set of all couples (ψ,φ) of Borel measurable bounded functions on $\mathcal{X}$ such that $\psi(x)+\varphi(y)\le c(x,y)$ for all $x,y\in\mathcal{X}$. This result is known as Kantorovich duality theorem and it holds true provided that c is lower semicontinuous. It still holds if Φ_c is replaced by $C_b\cap\Phi_c$ which is the subset of all couples $(\psi,\varphi)\in\Phi_c$ of continuous bounded functions. In the special case where $c=d$ is a lower semicontinuous metric, the above dual equality also holds with Φ_d the set of all couples (ψ,φ) of measurable (or continuous as well) bounded functions such that $\psi=-\varphi$ and φ is a d-Lipschitz function with a Lipschitz constant less than 1. In other words,

$$\mathcal{T}_d(\mu,\nu) = \sup\left\{ \int_{\mathcal{X}} \varphi\, d(\nu-\mu); \varphi\in B(\mathcal{X}), \|\varphi\|_{\mathrm{Lip}}\le 1 \right\} := \|\nu-\mu\|^*_{\mathrm{Lip}} \tag{2.2}$$

where the space of all Borel measurable bounded functions on $\mathcal{X}$ is denoted $B(\mathcal{X})$ and $\|\varphi\|_{\mathrm{Lip}} = \sup_{x\neq y} \frac{|\varphi(x)-\varphi(y)|}{d(x,y)}$ is the usual Lipschitz seminorm. This result, known as Kantorovich-Rubinstein's theorem, identifies the transportation cost $\mathcal{T}_d(\mu,\nu)$ with the dual norm $\|\nu-\mu\|^*_{\mathrm{Lip}}$.

2.1. **A larger class of transportation cost inequalities: $\mathcal{T}$-inequalities.**

After these considerations, it appears that the transportation cost inequality (1.1) enters the following larger class of inequalities, which we call $\mathcal{T}$-inequalities:

$$\alpha(\mathcal{T}(\nu)) \le H(\nu\mid\mu), \quad \forall\nu\in\mathcal{N} \tag{2.3}$$

where $\alpha:[0,\infty)\to[0,\infty)$ is an increasing function which vanishes at 0, $\mathcal{N}$ is a subset of $\mathcal{P}(\mathcal{X})$ and $\mathcal{T}$ is defined by

$$\mathcal{T}(\nu) = \sup_{(\psi,\varphi)\in\Phi} \left\{ \int_{\mathcal{X}} \psi\, d\mu + \int_{\mathcal{X}} \varphi\, d\nu \right\} \tag{2.4}$$

where Φ is a class of couples of functions (ψ, φ) with ψ integrable with respect to μ and φ integrable with respect to ν. Note that (2.3) is a family of inequalities where the value $+\infty$ is allowed with the convention that $\alpha(+\infty) = \lim_{t\to\infty} \alpha(t)$.
We are going to consider two cases which corresponds to what will be called Transportation Cost Inequalities and Norm-Entropy Inequalities.
Transportation Cost Inequalities. We assume that c is a nonnegative lower semicontinuous cost function. The space of all continous bounded functions on $\mathcal{X}$ is denoted $C_b(\mathcal{X})$. In the situation where Φ is equal to

$$\Phi_c := \{(\psi, \varphi) \in C_b(\mathcal{X}) \times C_b(\mathcal{X}); \psi \oplus \varphi \le c\}$$

the family of inequalities (2.3) is called a Transportation Cost Inequality (TCI). Indeed, the Kantorovich dual equality (2.1) states that

$$\mathcal{T}(\nu) = \mathcal{T}_c(\mu, \nu) \in [0, \infty],$$

for all $\nu \in \mathcal{N} \subset \mathcal{P}(\mathcal{X})$. In this situation, inequality (2.3) is

$$\alpha(\mathcal{T}_c(\mu, \nu)) \le H(\nu \mid \mu), \quad \forall \nu \in \mathcal{N} \tag{2.5}$$

Suppose that there exists a nonnegative measurable function χ on $\mathcal{X}$ such that $c(x, y) \le \chi(x) + \chi(y)$ for all $x, y \in \mathcal{X}$ and $\int_{\mathcal{X}} \chi\, d\mu < \infty$. A natural set $\mathcal{N}$ is the set of all probability measures ν such that $\int_{\mathcal{X}} \chi\, d\nu < \infty$.
Norm-Entropy Inequalities. Let U be a set of measurable functions on $\mathcal{X}$ such that $U = -U$. Let us take $\Phi = \Phi_U$ with

$$\Phi_U := \{(-\varphi, \varphi); \varphi \in U\}$$

This gives

$$\mathcal{T}(\nu) = \sup_{\varphi\in U} \int_{\mathcal{X}} \varphi\, d(\nu - \mu) := \|\nu - \mu\|^*_U \in [0, \infty].$$

In this case, inequality (2.3) is

$$\alpha(\|\nu - \mu\|^*_U) \le H(\nu \mid \mu), \quad \forall \nu \in \mathcal{P}_U \tag{2.6}$$

where $\mathcal{P}_U$ is the set of all $\nu \in \mathcal{P}(\mathcal{X})$ such that $\int_{\mathcal{X}} |\varphi|\, d\nu < \infty$ for all $\varphi \in U$. The family of inequalities (2.6) is called a Norm-Entropy Inequality (NEI).
As a typical example, let $(F, \|\cdot\|)$ be a seminormed space of measurable functions on $\mathcal{X}$ and $U := \{\varphi \in F, \|\varphi\| \le 1\}$ its unit ball. Then, $\|\nu - \mu\|^*_U$ is the dual norm of $\|\cdot\|$.

In the case where the cost function of a TCI is a lower semicontinuous metric d, the Kantorovich-Rubinstein theorem (see (2.2)) states that

$$\mathcal{T}_d(\mu, \nu) = \|\nu - \mu\|^*_{\mathrm{Lip}}$$

for all $\mu, \nu \in \mathcal{P}(\mathcal{X})$, where Φ_U is built with F the space all bounded d-Lipschitz functions on $\mathcal{X}$ endowed with the seminorm $\|\cdot\|_{\mathrm{Lip}}$. In this special important case, TCI and NEI match.

2.2. **Large deviations enter the game.** At Sections 3 and 7, $\mathcal{T}$-inequalities will be proved by means of a large deviation approach. The integral functional $H(\cdot \mid \mu)$ will be interpreted as the rate function of the large deviation principle (LDP) of the sequence of the empirical measures

$$L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$$

of an iid sample (X_i) of the law μ (δ_x stands for the Dirac measure at x). Indeed, by Sanov's theorem $\{L_n\}$ obeys the LDP in $\mathcal{P}(\mathcal{X})$ with the rate function

$$I(\nu) := H(\nu \mid \mu),\ \nu \in \mathcal{N}.$$

Roughly speaking, the sequence of random variables $\{L_n\}$ obeys the LDP in $\mathcal{N}$ with the rate function I if one has the following collection of estimates

$$\mathbb{P}(L_n \in A) \asymp \exp[-n \inf_{\nu \in A} I(\nu)]$$

as n tends to infinity, for any A "good" subset of $\mathcal{N}$. Let us introduce the nonnegative random variables

$$T_n = \mathcal{T}(L_n), \quad n \geq 1.$$

Suppose that $\mathcal{T}$ is regular enough for the sets $A_t = \{\nu \in \mathcal{N}, \mathcal{T}(\nu) \geq t\}$, $t \geq 0$, to be "good" sets. This means that for all $t \geq 0$,

$$\mathbb{P}(T_n \geq t) = \mathbb{P}(L_n \in A_t) \asymp \exp[-ni(t)]$$

with $i(t) = \inf\{I(\nu), \nu \in \mathcal{N}, \mathcal{T}(\nu) \geq t\} \in [0, \infty]$. Suppose that α is a *deviation function* for the sequence $\{T_n\}$ in the sense that it is an increasing nonnegative function on $[0, \infty)$ such that for all $t \geq 0$

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T_n \geq t) \leq -\alpha(t). \tag{2.7}$$

We obtain $\alpha(t) \leq i(t)$ for all t and in particular with $t = \mathcal{T}(\nu)$, we obtain for all $\nu \in \mathcal{N}$, $\alpha(\mathcal{T}(\nu)) \leq i(\mathcal{T}(\nu)) \leq I(\nu)$. This is precisely the desired inequality (2.3).
The recipe is:

Recipe 2.8. Any deviation function α of $\{T_n\}$ satisfies the $\mathcal{T}$-inequality (2.3).

Because of the sup entering the definition of $T_n = \sup_\Phi(\langle \varphi, L_n\rangle + \langle \psi, \mu\rangle)$, one may expect to get into troubles when trying to prove a full LDP for $\{T_n\}$. Fortunately, only the subclass of "deviation sets" $A_t = \{\nu \in \mathcal{N}, \mathcal{T}(\nu) \geq t\}$, $t \geq 0$, will be really useful.
This line of reasoning will be put on a solid ground at Theorem 3.7, Proposition 5.5 and Theorem 7.1.

2.3. **An example: CKP inequality.** As a simple illustration, we propose to prove CKP inequality by searching a deviation function α in the sense of (2.7). This is not intended to be the shortest proof, but only an illustration of the proposed method. Recall that CKP inequality is

$$\frac{1}{2}\|\nu - \mu\|_{\mathrm{TV}}^2 \leq H(\nu \mid \mu),\ \forall \nu \in \mathcal{P}(\mathcal{X}) \tag{2.9}$$

where $\|\xi\|_{\mathrm{TV}}$ is the total variation of the signed bounded measure ξ. As

$$\|\xi\|_{\mathrm{TV}} = \sup\left\{\int_{\mathcal{X}} \varphi\, d\xi, ; \varphi \text{ measurable such that } \|\varphi\| := \sup_{x \in \mathcal{X}} |\varphi(x)| \leq 1\right\},$$

(2.9) is the NEI with $F = B(\mathcal{X})$ the space of bounded measurable functions furnished with the uniform norm $\|\varphi\| := \sup_{x\in\mathcal{X}} |\varphi(x)|$, $\mathcal{N} = \mathcal{P}(\mathcal{X})$ and $\alpha(t) = t^2/2$.
Consider an iid sample (X_i) of the law μ and its associated sequence of empirical measures $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. For all n and all $\varphi \in U = \{\varphi \in B(\mathcal{X}); \|\varphi\| \le 1\}$, define the random variable

$$T_n^\varphi = \langle \varphi, L_n - \mu\rangle = \frac{1}{n}\sum_{i=1}^n Y_i^\varphi$$

where $Y_i^\varphi = \varphi(X_i) - \mathbb{E}\varphi(X_i)$. Cramér's theorem states that $\{T_n^\varphi\}$ obeys the LDP in $\mathbb{R}$ with rate function Λ_φ^* : the convex conjugate of the log-Laplace transform $\Lambda_\varphi(s) = \log \mathbb{E}e^{sY^\varphi}$, $s \in \mathbb{R}$. Recall that the convex conjugate of f is defined by $f^*(t) = \sup_{s\in\mathbb{R}}\{st - f(s)\} \in (-\infty, \infty]$, $t \in \mathbb{R}$.
Sanov's theorem holds in $\mathcal{P}(\mathcal{X})$ with the weak topology $\sigma(\mathcal{P}(\mathcal{X}), B(\mathcal{X}))$. As, $\nu \in \mathcal{P}(\mathcal{X}) \mapsto \langle \varphi, \nu - \mu\rangle$ is $\sigma(\mathcal{P}(\mathcal{X}), B(\mathcal{X}))$-continuous for all $\varphi \in B(\mathcal{X})$, one can apply the contraction principle. It gives us for all t

$$\Lambda_\varphi^*(t) = \inf\{H(\nu \mid \mu); \nu \in \mathcal{P}(\mathcal{X}) : \langle \varphi, \nu - \mu\rangle = t\},$$

which in turn implies that for all $\varphi \in B(\mathcal{X})$,

$$\Lambda_\varphi^*(\langle \varphi, \nu - \mu\rangle) \le H(\nu \mid \mu),\ \forall \nu \in \mathcal{P}(\mathcal{X}).$$

As Y^φ takes its values in $[\mathbb{E}Y^\varphi - 1, \mathbb{E}Y^\varphi + 1]$, by Hoeffding's inequality we have

$$\Lambda_\varphi(s) \le s^2/2 \tag{2.10}$$

for all real s. It follows that $\Lambda_\varphi^*(t) \ge \sup_{s\in\mathbb{R}}\{st - s^2/2\} = t^2/2$ for all real t. Hence, we have proved that for all $\varphi \in U$,

$$\alpha(\langle \varphi, \nu - \mu\rangle) \le H(\nu \mid \mu),\ \forall \nu \in \mathcal{P}(\mathcal{X})$$

with $\alpha(t) = t^2/2$. It follows that $\alpha(\sup_{\varphi\in U}\langle \varphi, \nu - \mu\rangle) \le H(\nu \mid \mu)$ for all $\nu \in \mathcal{P}(\mathcal{X})$, which is CKP inequality (2.9).
Some comments. In this proof, something interesting occured. Let us denote $T_n := \sup_\varphi T_n^\varphi$, $\beta(t) = -\limsup_{n\to\infty} \frac{1}{n}\log \mathbb{P}(T_n \ge t)$ and $J_\varphi(t) = -\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}(T_n^\varphi \ge t)$ the deviation functions of T_n and T_n^φ. As $T_n \ge T_n^\varphi$ for all φ, we have $\beta \le \inf_\varphi J_\varphi$. This means that a priori $\inf_\varphi J_\varphi$ could be too large to be the α of the NEI. On the other hand, by (2.10): $\sup_\varphi \Lambda_\varphi(s) \le \Lambda(s) := s^2/2$ for all $s \ge 0$, so that $t^2/2 = \Lambda^*(t) \le \inf_\varphi J_\varphi(t)$. Nevertheless, we have shown that Λ^* is a convenient function α for our NEI.
It will shown in a more general setting, at Theorem 7.7, that the convex lower semicontinuous envelope of $\inf_\varphi J_\varphi$ is the best *increasing convex* function α for this NEI.

3. Convex $\mathcal{T}$-inequalities. A dual characterization

In the rest of the paper (except Section 7) our attention is restricted to those $\mathcal{T}$-inequalities (2.3) where the function α is increasing and convex. In this case, (2.3) is said to be a convex $\mathcal{T}$-inequality.

3.1. **Sanov's theorem.** This theorem will be central for the proof of the main result of this section which is stated at Theorem 3.7.
Let the probability measure μ on $\mathcal{X}$ be given. We consider a sequence of independent $\mathcal{X}$-valued random variables $(X_i)_{i\geq 1}$ identically distributed with law μ. For any n the empirical measure of this sample is

$$L_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i} \in \mathcal{P}(\mathcal{X}).$$

We introduce the function space

$$\mathcal{F}_{\exp}(\mu) = \left\{\varphi : \mathcal{X} \to \mathbb{R}; \varphi \text{ measurable, } \int_{\mathcal{X}} \exp(a|\varphi|)\, d\mu < \infty \text{ for all } a > 0\right\} \tag{3.1}$$

of all the functions which admit exponential moments of all orders with respect to the measure μ. We denote

$$\mathcal{N}_{\exp}(\mu) = \left\{\nu \in \mathcal{P}(\mathcal{X}); \int_{\mathcal{X}} |\varphi|\, d\nu < \infty \text{ for all } \varphi \in \mathcal{F}_{\exp}(\mu)\right\}$$

the set of all probability measures which integrate every function of $\mathcal{F}_{\exp}(\mu)$.
The set $\mathcal{P}(\mathcal{X})$ is furnished with the cylinder σ-field generated by the functions $\nu \mapsto \langle \varphi, \nu\rangle$, $\varphi \in \mathcal{F}_{\exp}(\mu)$.

Theorem 3.2 (A version of Sanov's theorem)**.** *The effective domain of $H(\cdot \mid \mu)$ is included in $\mathcal{N}_{\exp}(\mu)$ and the sequence $\{L_n\}$ obeys the large deviation principle with rate function $H(\cdot \mid \mu)$ in $\mathcal{N}_{\exp}(\mu)$ equipped with the weak topology $\sigma(\mathcal{N}_{\exp}(\mu), \mathcal{F}_{\exp}(\mu))$.*
This means that for all measurable subset A of $\mathcal{N}_{\exp}(\mu)$, we have

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(L_n \in A) &\geq -\inf_{\nu \in \mathrm{int}\ A} H(\nu \mid \mu) \quad \textit{and} \\
\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(L_n \in A) &\leq -\inf_{\nu \in \mathrm{cl}\ A} H(\nu \mid \mu)
\end{aligned}$$

where int A *and* cl A *are the interior and closure of A.*

Proof. The proof is a variation of the classical proof of Sanov's theorem based on projective limits of LD systems (see [6], Thm 6.2.10). For two distinct detailed proofs of the present theorem, see ([8], Theorem 1.7) or ([11], Corollary 3.3). □

3.2. **The class of functions $\mathcal{C}$.** The functions α to be considered are assumed to be convex. Since α is also left continuous and increasing, we consider the following class of functions.

Definition 3.3 (of $\mathcal{C}$)**.** The class $\mathcal{C}$ consists of all the functions α on $[0, \infty)$ which are *convex* increasing, left continuous with $\alpha(0) = 0$.

For any α belonging to the class $\mathcal{C}$, denoting $t_* = \sup\{t \geq 0; \alpha(t) < \infty\}$, α is continuous on $[0, t_*)$ and $\lim_{t\uparrow t_*} \alpha(t) = \alpha(t_*)$.
The convex conjugate of a function $\alpha \in \mathcal{C}$ is replaced by the monotone conjugate $\alpha^{\circledast}$ defined by

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

where the supremum in taken on $t \geq 0$ instead of $t \in \mathbb{R}$. In fact, if α is extended by $\widetilde{\alpha}(t) = \begin{cases} \alpha(t) & \text{if } t \geq 0 \\ 0 & \text{if } t \leq 0 \end{cases}$ then the usual convex conjugate of $\widetilde{\alpha}$ is $\widetilde{\alpha}^*(s) = \begin{cases} \alpha^{\circledast}(s) & \text{if } s \geq 0 \\ +\infty & \text{if } s < 0 \end{cases}$. As $\widetilde{\alpha}$ is convex and lower semicontinuous, we have $\widetilde{\alpha}^{**} = \widetilde{\alpha}$. From this, it is not hard to deduce the following result.

Proposition 3.4. *For any function α on $[0,\infty)$, we have*

(a) $\alpha \in \mathcal{C} \Leftrightarrow \alpha^{\circledast} \in \mathcal{C}$
(b) $\alpha \in \mathcal{C} \Rightarrow \alpha^{\circledast\circledast} = \alpha$.

3.3. A convex criterion.

Theorem 3.7 below is a criterion for a convex $\mathcal{T}$-inequality to hold. It extends two well-known results of S. Bobkov and F. Götze ([2], Theorem 1.3 and statement (1.7)).
Let $\mathcal{F}$ be a vector space of measurable functions φ on $\mathcal{X}$ such that

$$\int_{\mathcal{X}} e^{\varphi}\, d\mu < \infty, \quad \forall \varphi \in \mathcal{F}. \tag{3.5}$$

Let $\mathcal{P}_{\mathcal{F}}$ be the set of all probability measures which integrate $\mathcal{F}$:

$$\mathcal{P}_{\mathcal{F}} = \left\{ \nu \in \mathcal{P}(\mathcal{X}); \int_{\mathcal{X}} |\varphi|\, d\nu < \infty,\ \forall \varphi \in \mathcal{F} \right\}.$$

Clearly, if the class Φ entering the definition of $\mathcal{T}(\nu)$ satisfies

$$(0,0) \in \Phi \subset \mathcal{F} \times \mathcal{F}, \tag{3.6}$$

the function $\mathcal{T}$ is a well defined $[0,\infty]$-valued function on $\mathcal{P}_{\mathcal{F}}$.
Let $\Lambda_\phi(s)$ be the log-Laplace transform of $\varphi(X) + \mathbb{E}\psi(X)$ where X admits μ as its law. We have for all real s,

$$\Lambda_\phi(s) = \log \int_{\mathcal{X}} \exp[s(\varphi(x) + \langle \psi, \mu \rangle)]\, \mu(dx)$$

Theorem 3.7. *We assume (3.5) and (3.6). Let us consider the following statements where α is any function in $\mathcal{C}$:*

(a) $\alpha(\mathcal{T}(\nu)) \leq H(\nu \mid \mu), \quad \forall \nu \in \mathcal{P}_{\mathcal{F}}$.
(b) $\Lambda_\phi(s) \leq \alpha^{\circledast}(s), \quad \forall s \geq 0,\ \forall \phi \in \Phi$.
(c) $\alpha(t) \leq \Lambda_\phi^*(t), \forall t \geq 0, \forall \phi \in \Phi$.
(d) $\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\langle \varphi, L_n \rangle + \langle \psi, \mu \rangle \geq t) \leq -\alpha(t), \quad \forall t \geq 0,\ \forall (\psi,\varphi) \in \Phi$.
(e) $\forall n \geq 1,\ \frac{1}{n} \log \mathbb{P}(\langle \varphi, L_n \rangle + \langle \psi, \mu \rangle \geq t) \leq -\alpha(t), \quad \forall t \geq 0,\ \forall (\psi,\varphi) \in \Phi$.

Then, we have (a) $\Leftrightarrow$ (b) $\Leftrightarrow$ (c) and (e) $\Rightarrow$ (d) $\Rightarrow$ (a).
If it is assumed in addition that for all $(\psi,\varphi) \in \Phi$,

$$\int_{\mathcal{X}} (\varphi(x) + \psi(x))\, \mu(dx) \leq 0 \tag{3.8}$$

then, we have (a) $\Leftrightarrow$ (b) $\Leftrightarrow$ (c) $\Leftrightarrow$ (d) $\Leftrightarrow$ (e).

The most useful statement of this theorem is the criterion $(b) \Rightarrow (a)$.
Clearly, the requirement (3.8) holds for all NEIs. It also holds for TCIs under the assumption that c satisfies

$$c(x,x) = 0, \ \forall x \in \mathcal{X}. \tag{3.9}$$

When working with TCIs, this will be assumed in the sequel.

Proof. Possibly considering the vector space $\mathcal{F}'$ spanned by $\mathcal{F} \cup C_b(\mathcal{X})$ instead of $\mathcal{F}$, one can assume that $\mathcal{F}$ separates $\mathcal{P}_{\mathcal{F}}$. Indeed, the assumptions (3.5) and (3.6) still hold with $\mathcal{F}'$ instead of $\mathcal{F}$ and we clearly have $\mathcal{P}_{\mathcal{F}'} = \mathcal{P}_{\mathcal{F}}$. Hence, we assume without loss of generality that $\mathcal{F}$ separates $\mathcal{P}_{\mathcal{F}}$. As a consequence, the weak topology $\sigma(\mathcal{P}_{\mathcal{F}}, \mathcal{F})$ is Hausdorff: this is necessary to derive LDPs away from compactness troubles.
Note that the assumption (3.5) is equivalent to $\mathcal{F} \subset \mathcal{F}_{\exp}(\mu)$. It follows that under this assumption, Sanov's Theorem 3.2 implies that $\{L_n\}$ obeys the LDP in $\mathcal{P}_{\mathcal{F}}$ equipped with $\sigma(\mathcal{P}_{\mathcal{F}}, \mathcal{F})$ with $H(\cdot \mid \mu)$ as its rate function.
Consider, for any $(\psi, \varphi) := \phi \in \Phi$ and $n \geq 1$,

$$T_n^\phi = \langle \varphi, L_n \rangle + \langle \psi, \mu \rangle = \frac{1}{n} \sum_{i=1}^n (\varphi(X_i) + \mathbb{E}\psi(X_i)) \tag{3.10}$$

so that $T_n := \mathcal{T}(L_n) = \sup_{\phi \in \Phi} T_n^\phi$. Cramér's theorem states that $\{T_n^\phi\}$ obeys the LDP in $\mathbb{R}$ with

$$\Lambda_\phi^*(t) = \sup_{s \in \mathbb{R}} \{st - \Lambda_\phi(s)\}, \ t \in \mathbb{R}$$

as its rate function. In particular, for all real t

$$\begin{aligned} -\inf_{u > t} \Lambda_\phi^*(u) &\leq \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(T_n^\phi > t) \\ &\leq \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(T_n^\phi \geq t) \leq -\inf_{u \geq t} \Lambda_\phi^*(u) \end{aligned} \tag{3.11}$$

Because of assumption (3.6), the mapping $f_\phi : \nu \in \mathcal{P}_{\mathcal{F}} \mapsto \langle \varphi, \nu \rangle + \langle \psi, \mu \rangle \in \mathbb{R}$ is continuous for every $(\psi, \varphi) \in \Phi$. As $T_n^\phi = f_\phi(L_n)$, one can apply the contraction principle which gives us for all real t

$$\Lambda_\phi^*(t) = \inf\{H(\nu \mid \mu); \nu \in \mathcal{P}_{\mathcal{F}} : \langle \varphi, \nu \rangle + \langle \psi, \mu \rangle = t\}. \tag{3.12}$$

$[(a) \Leftrightarrow (c)]$:

$$\begin{aligned} (a) &\overset{(i)}{\Leftrightarrow} \alpha\left(\sup_\phi(\langle \varphi, \nu \rangle + \langle \psi, \mu \rangle)\right) \leq H(\nu \mid \mu), \forall \nu \in \mathcal{P}_{\mathcal{F}} \\ &\overset{(ii)}{\Leftrightarrow} \alpha(\langle \varphi, \nu \rangle + \langle \psi, \mu \rangle) \leq H(\nu \mid \mu), \forall \nu \in \mathcal{P}_{\mathcal{F}}, \forall \phi \in \Phi \\ &\Leftrightarrow \alpha(t) \leq H(\nu \mid \mu), \forall t \in \mathbb{R}, \forall \phi \in \Phi, \forall \nu \in \mathcal{P}_{\mathcal{F}} : \langle \varphi, \nu \rangle + \langle \psi, \mu \rangle = t \\ &\Leftrightarrow \alpha(t) \leq \inf\{H(\nu \mid \mu); \nu \in \mathcal{P}_{\mathcal{F}} : \langle \varphi, \nu \rangle + \langle \psi, \mu \rangle = t\}, \forall t \in \mathbb{R}, \forall \phi \in \Phi \\ &\overset{(iii)}{\Leftrightarrow} \alpha \leq \Lambda_\phi^* \\ &\Leftrightarrow (c) \end{aligned}$$

The equivalence (i) follows from the definition (2.4) of $\mathcal{T}$, (ii) holds true because α is increasing and left continuous while (iii) follows from (3.12).

$[(b) \Leftrightarrow (c)]$. In order to work with usual convex conjugates instead of monotone conjugates, let us take $\alpha^*(s) = +\infty$ for all $s < 0$. It follows that α is extended by $\alpha(t) = 0$, for all $t \leq 0$ and $\alpha^{\circledast}(s) = \alpha^*(s)$ for all $s \geq 0$.
Let us prove $(c) \Rightarrow (b)$. With the above convention, statement (c) is equivalent to

$$\alpha(t) \leq \Lambda_\phi^*(t), \forall t \in \mathbb{R}, \forall \phi \in \Phi. \tag{3.13}$$

As, Λ_ϕ is convex and lower semicontinuous, we have: $\Lambda_\phi^{**} = \Lambda_\phi$. Hence, taking the convex conjugates on both sides of (3.13) one obtains that $\Lambda_\phi \leq \alpha^*$ which entails (b).
Let us prove $(b) \Rightarrow (c)$. As α is in $\mathcal{C}$, its extension (still denoted by α) is convex and lower semicontinuous, so that $\alpha^{**} = \alpha$. Therefore, taking the conjugate of (b) leads to $\alpha \leq \Lambda_\phi^*$ which is (c).
The convexity of α has been used to obtain $(b) \Rightarrow (c)$ and it won't be used anywhere else.

$[(e) \Rightarrow (d) \Rightarrow (a)]$. As $(e) \Rightarrow (d)$ is obvious and $(a) \Leftrightarrow (c)$, all we have to show is $(d) \Rightarrow (c)$.
Let $m = \mathbb{E}Y = \langle \varphi + \psi, \mu \rangle$. For all $t \leq m$, we have $\inf_{u>t} \Lambda_\phi^*(u) = \inf_{u \geq t} \Lambda_\phi^*(u) = 0$. As Λ_ϕ^* is convex, it is continuous on (t_-, t_+) the interior of its effective domain. Therefore, we have for all $t \neq t_+$, $\inf_{u>t} \Lambda_\phi^*(u) = \inf_{u \geq t} \Lambda_\phi^*(u)$. Together with (3.11), this gives for all $t \neq t_+$,

$$-\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T_n^\phi \geq t) = \inf_{u>t} \Lambda_\phi^*(u) = \inf_{u \geq t} \Lambda_\phi^*(u) = \begin{cases} 0, & \text{if } t \leq m \\ \Lambda_\phi^*(t), & \text{if } t \geq m \end{cases} = \Lambda_\phi^{\circledast}(t).$$

Consequently, considering $\Gamma(t) = \Lambda_\phi^{\circledast}(t)$ if $t \neq t_+$ and $\Gamma(t_+) = +\infty$ (if $t_+ < \infty$), we have

$$\begin{aligned} (d) \quad &\Rightarrow \quad \alpha(t) \leq \Lambda_\phi^{\circledast}(t), \ \forall t \neq t_+ \\ &\Rightarrow \quad \alpha \leq \Gamma \\ &\Rightarrow \quad \mathrm{ls}\ \alpha \leq \mathrm{ls}\ \Gamma \\ &\Rightarrow \quad \alpha \leq \Lambda_\phi^{\circledast} \end{aligned}$$

where $\mathrm{ls}\ \alpha$ and $\mathrm{ls}\ \Gamma$ are the lower semicontinuous envelopes of α and Γ, and the last implication holds since α is lower semicontinuous and $\mathrm{ls}\ \Gamma = \Lambda_\phi^{\circledast}$. As $\Lambda_\phi^{\circledast} \leq \Lambda_\phi^*$, we have the desired result.

$[(a) \Leftrightarrow (b) \Leftrightarrow (c) \Leftrightarrow (d) \Leftrightarrow (e)]$. Let us assume (3.8). To obtain the stated series of equivalences, it remains to prove $(c) \Rightarrow (e)$.
By (3.10), $T_n^\phi = \frac{1}{n} \sum_{i=1}^n Y_i$ with $Y_i = \varphi(X_i) + \mathbb{E}\psi(X_i)$. The standard proof of the upper bound of Cramér's theorem is based on an optimization of a collection of exponential Markov inequalities, as follows. For all real t, all n and all $s \geq 0$,

$$\begin{aligned} \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^n Y_i \geq t\right) \quad &\leq \quad \mathbb{P}\left(\exp[s\sum_{i=1}^n Y_i] \geq e^{nst}\right) \\ &\leq \quad e^{-nst}\mathbb{E}\exp[s\sum_{i=1}^n Y_i] \\ &= \quad \exp[n(\Lambda_\phi(s) - st)] \end{aligned}$$

Optimizing on $s \geq 0$, one obtains that

$$\frac{1}{n} \log \mathbb{P}(T_n^{\phi} \geq t) \leq -\Lambda_{\phi}^{\circledast}(t), \ \forall t \in \mathbb{R}, \forall \phi \in \Phi, \forall n \geq 1.$$

But, assumption (3.8) implies that $m \leq 0$ so that $\Lambda_{\phi}^{\circledast}(t) = \Lambda_{\phi}^{*}(t)$ for all $t \geq 0$. It follows immediately that $(c) \Rightarrow (e)$. This completes the proof of the theorem. □

3.4. **Convex Transportation Cost Inequalities.** In the special case of TCIs, we have $\Phi = \Phi_c = \{(\psi, \varphi); \psi, \varphi \in C_b(\mathcal{X}) : \psi \oplus \varphi \leq c\}$. Optimal transportation theory (see [17]) indicates that Φ_c may be replaced with the smaller sets $\{(-\varphi, Q^c\varphi); \varphi \in C_b(\mathcal{X})\}$ or $\{(-\varphi, Q^c\varphi); \varphi$ lower semicontinuous and bounded on $\mathcal{X}\}$ where

$$Q^c\varphi(y) = \inf_{x \in \mathcal{X}} \{\varphi(x) + c(x, y)\}, \ y \in \mathcal{X}$$

without any change in the value of $\mathcal{T}_c$. One easily proves that if (3.9) is satisfied: $c(x, x) = 0$ for all $x \in \mathcal{X}$, then $\sup |Q^c\varphi| \leq \sup |\varphi|$. If c is continuous, then $Q^c\varphi$ is measurable as an upper semicontinuous function. If c is only assumed to be lower semicontinuous, $Q^c\varphi$ is still measurable if φ is lower semicontinuous and bounded (but the proof of this result is technical). Anyway, $Q^c\varphi \in B(\mathcal{X})$ (is a bounded measurable function) as soon as φ is lower semicontinuous and bounded. In particular, assumptions (3.5) and (3.6) hold with $\mathcal{F} = B(\mathcal{X})$.
Now, as a corollary of Theorem 3.7, we have the following result.

Corollary 3.14. *Whenever $\alpha \in \mathcal{C}$, the transportation cost inequality (2.5) holds in $\mathcal{N} = \mathcal{P}(\mathcal{X})$ if and only if*

$$\log \int_{\mathcal{X}} e^{s[Q^c\varphi(y) - \langle \varphi, \mu \rangle]} \, \mu(dy) \leq \alpha^{\circledast}(s)$$

for all $s \geq 0$ and all $\varphi \in C_b(\mathcal{X})$.
If in addition c is continuous, the same result holds when $\varphi \in C_b(\mathcal{X})$ is replaced with $\varphi \in B(\mathcal{X})$: the set of all measurable bounded functions on $\mathcal{X}$.

3.5. **Convex Norm-Entropy inequalities.** In the special case of NEIs, we have $\Phi = \{(-\varphi, \varphi); \varphi \in U\}$ and Theorem 3.7 specializes as follows.

Theorem 3.15. *Suppose that U satisfies*

$$\int_{\mathcal{X}} e^{a|\varphi|} \, d\mu < \infty, \forall \varphi \in U, \forall a > 0.$$

Let α be in $\mathcal{C}$. Then, the norm-entropy inequality (2.6)

$$\alpha(\|\nu - \mu\|_U^*) \leq H(\nu \mid \mu), \quad \forall \nu \in \mathcal{P}_U$$

holds if and only if

$$\Lambda_{\varphi}(s) := \log \int_{\mathcal{X}} e^{s[\varphi(x) - \langle \varphi, \mu \rangle]} \, \mu(dx) \leq \alpha^{\circledast}(s) \tag{3.16}$$

for all $s \geq 0$ and all $\varphi \in U$.

Specializing Theorem 3.15 by taking U to be the set of all 1-Lipschitz measurable bounded functions with respect some measurable metric d, one obtains the following characterization of convex T_1-inequalities.

Theorem 3.17 (T_1-inequality)**.** *Let d be a lower semicontinuous metric on $\mathcal{X}$ such that*
$$\int_{\mathcal{X}} e^{a_o d(x_o,x)}\,\mu(dx)<\infty,$$
for some $a_o>0$ and some (and therefore all) $x_o\in\mathcal{X}$. Let α be in $\mathcal{C}$. Then,
$$\alpha(\mathcal{T}_d(\mu,\nu))\le H(\nu\mid\mu),$$
for all $\nu\in\mathcal{P}(\mathcal{X})$ such that $\int_{\mathcal{X}} d(x_o,x)\,\nu(dx)<\infty$ if and only if
$$\Lambda_\varphi(s):=\log\int_{\mathcal{X}} e^{s[\varphi(x)-\langle\varphi,\mu\rangle]}\,\mu(dx)\le\alpha^{\circledast}(s) \tag{3.18}$$
for all $s\ge0$ and all measurable bounded Lipschitz function φ such that $\|\varphi\|_{\mathrm{Lip}}\le1$.

The following simple result asserts that the functions α of NEIs cannot grow faster than at^2 for t near zero.

Proposition 3.19. *Assuming that F contains functions which are not μ-a.e. constant, the function α of a convex norm-entropy inequality (2.6) satisfies*
$$0\le\alpha(t)\le at^2,\forall 0\le t\le t_1 \tag{3.20}$$
for some $a>0$ and $t_1>0$.

Proof. Let φ_o be a non constant function in U. Then, $\sigma_o^2:=\int_{\mathcal{X}}(\varphi(x)-\langle\varphi,\mu\rangle)^2\,d\mu>0$ and for any $0<\sigma_1^2<\sigma_o^2$ there exists $s_1>0$ such that $\Lambda_{\varphi_o}(s)=\sigma_o^2s^2/2+o(s^2)\ge\sigma_1^2s^2/2$, for all $0\le s\le s_1$. Let $\theta_1(s)$ match with $\sigma_1^2s^2/2$ on $[0,s_1]$ and be extended on $[s_1,\infty)$ by the tangent affine function of $s\mapsto\sigma_1^2s^2/2$ at $s=s_1$. As Λ_{φ_o} is convex, we have $\theta_1(s)\le\Lambda_{\varphi_o}(s)$ for all $s\ge0$.
Together with (3.16), we obtain $\theta_1\le\alpha^{\circledast}$. Taking the monotone conjugates on both sides of this inequality provides us with
$$\alpha(t)\le\theta_1^{\circledast}(t)=\begin{cases} t^2/(2\sigma_1^2), & \text{if } 0\le t\le s_1\sigma_1^2\\ +\infty, & \text{if } t>s_1\sigma_1^2\end{cases}$$
from which the desired result follows. □

To explore some consequences of Theorem 3.15 (see Corollaries 3.23 and 3.24 below) one needs the notion of Orlicz space associated with the exponential function. It appears that the space $\mathcal{F}_{\exp}(\mu)$ introduced at (3.1) is the Orlicz space
$$\left\{\varphi:\mathcal{X}\to\mathbb{R};\ \text{measurable},\ \int_{\mathcal{X}}\rho(a\varphi)\,d\mu<\infty\text{ for all }a>0\right\}$$
where μ-almost equal functions are not identified and ρ is the Young function
$$\rho(s)=e^{|s|}-1,\quad s\in\mathbb{R}.$$
Its Orlicz norm is defined by
$$\begin{aligned}\|\varphi\|_\rho &:= \inf\left\{b>0;\int_{\mathcal{X}}\rho\left(\frac{\varphi}{b}\right)\,d\mu\le1\right\} \\ &= \inf\left\{b>0;\int_{\mathcal{X}}e^{|\varphi|/b}\,d\mu\le2\right\}\end{aligned} \tag{3.21}$$

and considering the usual dual bracket $\langle \eta, \varphi \rangle = \int_{\mathcal{X}} \eta\varphi \, d\mu$, its topological dual space is isomorphic to

$$\begin{aligned} L_{\rho^*}(\mu) &= \left\{ \eta : \mathcal{X} \to \mathbb{R};\ \text{measurable},\ \int_{\mathcal{X}} \rho^*(a\eta)\, d\mu < \infty \text{ for some } a > 0 \right\} \\ &= \left\{ \eta : \mathcal{X} \to \mathbb{R};\ \text{measurable},\ \int_{\mathcal{X}} |\eta| \log |\eta| \, d\mu < \infty \right\} \end{aligned}$$

where ρ^* is the convex conjugate of ρ :

$$\rho^*(t) = \begin{cases} |t| \log |t| - |t| + 1, & \text{if } |t| \geq 1 \\ 0, & \text{if } |t| \leq 1 \end{cases}$$

and μ-almost equal functions are identified. Note that the effective domain of $H(\cdot \mid \mu)$ is included in the set of all probability measures ν which are absolutely continuous with respect to μ and such that $\frac{d\nu}{d\mu} \in L_{\rho^*}(\mu)$.
Let us state a useful technical lemma, which will play a role that is similar to the role that Hoeffding's inequality (2.10) played during the proof of CKP inequality.

Lemma 3.22 (A Bernstein type inequality)**.** *For any measurable function φ such that $\int_{\mathcal{X}} e^{a_o|\varphi|} \, d\mu < \infty$ for some $a_o > 0$, we have $\|\varphi\|_\rho < \infty$ and*

$$\Lambda_\varphi(s) \leq \frac{\|\varphi\|_\rho^2 \, s^2}{1 - \|\varphi\|_\rho \, s}, \quad \forall\ 0 \leq s < 1/\|\varphi\|_\rho.$$

It follows that, if U is a uniformfy $\|\cdot\|_\rho$-bounded set of functions: $\sup_{\varphi \in U} \|\varphi\|_\rho \leq M < \infty$, *then*

$$\Lambda_\varphi(s) \leq \frac{M^2 s^2}{1 - Ms}, \quad \forall\ 0 \leq s < 1/M, \forall \varphi \in U.$$

Proof. By the definition of $\beta := \|\varphi\|_\rho$, we have $1 \geq \int_{\mathcal{X}} \rho(\varphi/\beta) \, d\mu = \sum_{k \geq 1} \langle |\varphi|^k, \mu \rangle / (k!\beta^k)$. Therefore, for all $k \geq 1$, $\langle |\varphi|^k, \mu \rangle \leq k!\beta^k$. It follows that for all $s \geq 0$,

$$\begin{aligned} \Lambda_\varphi(s) &= \log \left(1 + \sum_{k \geq 1} s^k \langle \varphi^k, \mu \rangle / k! \right) - s\langle \varphi, \mu \rangle \\ &\leq \sum_{k \geq 2} s^k \langle \varphi^k, \mu \rangle / k! \\ &\leq \sum_{k \geq 2} s^k \langle |\varphi|^k, \mu \rangle / k! \\ &\leq \sum_{k \geq 2} (\beta s)^k \\ &= \begin{cases} (\beta s)^2/(1 - \beta s), & \text{if } 0 \leq \beta s < 1 \\ +\infty, & \text{if } \beta s \geq 1 \end{cases} \end{aligned}$$

The last statement holds since $\beta \mapsto \sum_{k \geq 2} (\beta s)^k$ is an increasing function, for all $s \geq 0$. $\square$

We are now ready to prove some corollaries of Theorem 3.7.

For any measurable function f in $L_{\rho^*}(\mu)$, let

$$\begin{aligned}\|f\|_\rho^* &:= \sup\left\{\int_{\mathcal{X}} f\varphi\,d\mu; \varphi : \text{measurable}, \ \|\varphi\|_\rho \le 1\right\}\\ &= \sup\left\{\int_{\mathcal{X}} f\varphi\,d\mu; \varphi : \text{measurable}, \ \int_{\mathcal{X}} e^{|\varphi|}\,d\mu \le 2\right\}\end{aligned}$$

be the dual norm of $\|\cdot\|_\rho$.

Corollary 3.23. *For any probability measure ν which is absolutely continuous with respect to μ and such that $\frac{d\nu}{d\mu} \in L_{\rho^*}(\mu)$, we have*

$$\left\|\frac{d\nu}{d\mu} - 1\right\|_\rho^* \le 2\sqrt{H(\nu \mid \mu)} + H(\nu \mid \mu).$$

Note that this is the NEI: $\alpha_1(\|\frac{d\nu}{d\mu} - 1\|_\rho^*) \le H(\nu \mid \mu)$, with $\alpha_1(t) = (\sqrt{t+1} - 1)^2$.

Proof. Here U is the unit ball of $\mathcal{F}_{\exp}(\mu)$ and thanks to Lemma 3.22 applied with $M = 1$, (3.16) holds as follows: $\Lambda_\varphi(s) \le \alpha_1^{\circledast}(s) := s^2/(1-s)$. Taking the monotone conjugate, we obtain $\alpha_1(t) = (\sqrt{t+1} - 1)^2$, which is the desired result. □

The following corollary has already been obtained by F. Bolley and C. Villani in [?] with other constants.

Corollary 3.24 (Weighted CKP inequalities)**.** *Let χ be a nonnegative function such that $\int_{\mathcal{X}} e^{a_o\chi}\,d\mu < \infty$ for some $a_o > 0$. Then, $\|\chi\|_\rho < \infty$ and for any probability measure ν which is absolutely continuous with respect to μ and such that $\frac{d\nu}{d\mu} \in L_{\rho^*}(\mu)$, $\|\chi\cdot(\nu-\mu)\|_{\mathrm{TV}}$ is well defined, finite and we have*

$$\|\chi\cdot(\nu-\mu)\|_{\mathrm{TV}} \le \|\chi\|_\rho \left(2\sqrt{H(\nu \mid \mu)} + H(\nu \mid \mu)\right)$$

Note that this is the NEI: $\alpha(\|\chi\cdot(\nu-\mu)\|_{\mathrm{TV}}) \le H(\nu \mid \mu)$, with $\alpha(t) = (\sqrt{t/\|\chi\|_\rho + 1} - 1)^2$.

Proof. Here $U = \{\chi\psi; \sup|\psi| \le 1\}$. As χ may not be in $\mathcal{F}_{\exp}(\mu)$ (if there exists $a_1 > 0$ such that $\int_{\mathcal{X}} e^{a_1\chi}\,d\mu = \infty$), one must be careful. It happens that

$$\begin{aligned}\|\chi\cdot(\nu-\mu)\|_{\mathrm{TV}} &= \sup\left\{\int_{\mathcal{X}} \chi\psi\,d(\nu-\mu); \psi : \text{measurable}, \ \sup|\psi| \le 1\right\}\\ &= \sup\left\{\int_{\mathcal{X}} \varphi\,d(\nu-\mu); \varphi : \text{measurable}, \ |\varphi| \le \chi, \sup|\varphi| < \infty\right\}.\end{aligned}$$

To show this, decompose $\nu - \mu$ into its positive and negative parts, approximate from below $\chi|\psi|\mathbf{1}_{\mathrm{supp}((\nu-\mu)_+)}$ and $\chi|\psi|\mathbf{1}_{\mathrm{supp}((\nu-\mu)_-)}$ by pointwise converging sequences of bounded functions, and conclude with the dominated convergence theorem.
Therefore, U can be replaced with $U' = \{\varphi; |\varphi| \le \chi, \sup|\varphi| < \infty\} \subset \mathcal{F}_{\exp}(\mu)$. As $\sup_{\varphi\in U'} \|\varphi\|_\rho \le \|\chi\|_\rho$, thanks to Lemma 3.22 applied with $M = \|\chi\|_\rho$, (3.16) holds as follows: $\Lambda_\varphi(s) \le \alpha_M^{\circledast}(s) := (Ms)^2/(1-Ms)$. Taking the monotone conjugate, we obtain $\alpha_M(t) = (\sqrt{t/M+1} - 1)^2$, which is the desired result. □

Remark 3.25. It follows from Corollaries 3.23 and 3.24, that

$$\|\nu-\mu\|_{\mathrm{TV}} \le \frac{1}{\log 2}\left(2\sqrt{H(\nu\mid\mu)}+H(\nu\mid\mu)\right),$$

which of course is worse than CKP inequality (2.9) but has the same order of growth $\sqrt{H}$ for vanishing entropies.

Let d be a metric on $\mathcal{X}$. The associated dual Lipschitz norm of any signed bounded measure ξ with *zero mass* is defined by

$$\|\xi\|^*_{\mathrm{Lip}} = \sup\left\{\int_{\mathcal{X}}\varphi\,d\xi; \varphi : \text{measurable}, \|\varphi\|_{\mathrm{Lip}} \le 1, \sup|\varphi| < \infty\right\}$$

where $\|\varphi\|_{\mathrm{Lip}} = \sup_{x\neq y}\frac{|\varphi(x)-\varphi(y)|}{d(x,y)}$ is the usual Lipschitz seminorm.

Corollary 3.26. *Suppose that there exist $a_o > 0$ and $x_o \in \mathcal{X}$ such that $\int_{\mathcal{X}} e^{a_o d(x_o,x)}\,\mu(dx) < \infty$. Then, $\|d\|_{\rho,\mu^{\otimes 2}} = \inf\{b>0; \int_{\mathcal{X}\times\mathcal{X}} e^{d(x,y)/b}\,\mu(dx)\mu(dy) \le 2\} < \infty$ and*

$$\|\nu-\mu\|^*_{\mathrm{Lip}} \le \|d\|_{\rho,\mu^{\otimes 2}}\left(2\sqrt{H(\nu\mid\mu)}+H(\nu\mid\mu)\right), \quad \forall\nu\in\mathcal{P}(\mathcal{X}).$$

Note that this is the NEI: $\alpha(\|\nu-\mu\|^*_{\mathrm{Lip}}) \le H(\nu\mid\mu)$, with $\alpha(t) = (\sqrt{t/\|d\|_{\rho,\mu^{\otimes 2}}+1}-1)^2$.

Proof. This is a corollary of Theorem 3.17. Here $U = \{\varphi : \|\varphi\|_{\mathrm{Lip}} \le 1, \sup|\varphi| < \infty\} \subset \mathcal{F}_{\exp}(\mu)$. Let us show that

$$\sup_{\varphi\in U}\|\varphi-\langle\varphi,\mu\rangle\|_\rho \le \|d\|_{\rho,\mu^{\otimes 2}}. \tag{3.27}$$

By Jensen's inequality, for any 1-Lipschitz function φ and all $s \ge 0$,

$$\begin{aligned}\exp\left[s\left(\varphi(x)-\int_{\mathcal{X}}\varphi(y)\,\mu(dy)\right)\right] &\le \int_{\mathcal{X}}\exp[s(\varphi(x)-\varphi(y))]\,\mu(dy)\\ &\le \int_{\mathcal{X}}\exp[sd(x,y)]\,\mu(dy).\end{aligned}$$

Hence, integrating with respect to $\mu(dx)$, one obtains (3.27).
Thanks to Lemma 3.22 applied with $M = \|d\|_{\rho,\mu^{\otimes 2}}$, (3.18) holds as follows: $\Lambda_\varphi(s) \le \alpha^{\circledast}_M(s) := (Ms)^2/(1-Ms)$. Taking the monotone conjugate, we obtain $\alpha_M(t) = (\sqrt{t/M+1}-1)^2$, which is the desired result. □

4. Tensorization of convex TCIs

In this section only convex TCIs are considered. It is assumed that the appearing state spaces are Polish and the appearing cost functions are nonnegative *continuous* and satisfy (3.9).

4.1. **Statement of the main result.** Let μ_1, μ_2 be two probability measures on two Polish spaces $\mathcal{X}_1$, $\mathcal{X}_2$, respectively. The cost functions $c_1(x_1,y_1)$ and $c_2(x_2,y_2)$ on $\mathcal{X}_1\times\mathcal{X}_1$ and $\mathcal{X}_2\times\mathcal{X}_2$ give rise to the optimal transportation cost functions $\mathcal{T}_{c_1}(\mu_1,\nu_1)$, $\nu_1\in\mathcal{P}(\mathcal{X}_1)$ and $\mathcal{T}_{c_2}(\mu_2,\nu_2)$, $\nu_2\in\mathcal{P}(\mathcal{X}_2)$.
On the product space $\mathcal{X}_1\times\mathcal{X}_2$, we now consider the product measure $\mu_1\otimes\mu_2$ and the cost function

$$c_1\oplus c_2\big((x_1,y_1),(x_2,y_2)\big) := c_1(x_1,y_1)+c_2(x_2,y_2),\quad x_1,y_1\in\mathcal{X}_1, x_2,y_2\in\mathcal{X}_2$$

which give rise to the so-called tensorized optimal transportation cost function

$$\mathcal{T}_{c_1\oplus c_2}(\mu_1\otimes\mu_2,\nu),\quad \nu\in\mathcal{P}(\mathcal{X}_1\times\mathcal{X}_2).$$

Recall that the inf-convolution of two functions α_1 and α_2 on $[0,\infty)$ is defined by

$$\alpha_1\Box\alpha_2(t)=\inf\{\alpha_1(t_1)+\alpha_2(t_2); t_1,t_2\geq 0: t_1+t_2=t\},\quad t\geq 0.$$

Lemma 4.1. *Let α_1 and α_2 belong to the class $\mathcal{C}$. Then,*
- (a) $\alpha_1\Box\alpha_2\in\mathcal{C}$ *and*
- (b) $(\alpha_1\Box\alpha_2)^{\circledast}=\alpha_1^{\circledast}+\alpha_2^{\circledast}$

Proof. This simple exercice is left to the reader. □

The main result of this section is the following theorem.

Theorem 4.2 (Tensorization)**.** *Let c_1 and c_2 be two continuous nonnegative cost functions which satisfy (3.9). Suppose that the convex TCIs*

$$\begin{aligned}\alpha_1(\mathcal{T}_{c_1}(\mu_1,\nu_1)) &\leq H(\nu_1\mid\mu_1),\ \forall\nu_1\in\mathcal{P}(\mathcal{X}_1)\\ \alpha_2(\mathcal{T}_{c_2}(\mu_2,\nu_2)) &\leq H(\nu_2\mid\mu_2),\ \forall\nu_2\in\mathcal{P}(\mathcal{X}_2)\end{aligned}$$

hold with $\alpha_1,\alpha_2\in\mathcal{C}$. Then, on the product space $\mathcal{X}_1\times\mathcal{X}_2$, we have the convex TCI

$$\alpha_1\Box\alpha_2\big(\mathcal{T}_{c_1\oplus c_2}(\mu_1\otimes\mu_2,\nu)\big)\leq H(\nu\mid\mu_1\otimes\mu_2),\quad\forall\nu\in\mathcal{P}(\mathcal{X}_1\times\mathcal{X}_2)$$

Its proof is postponed to Section 4.3. We prefer beginning with a presentation at the next section of an incomplete derivation of this result which, to our opinion, seems to be more intuitively appealing.

4.2. **An incomplete direct proof of Theorem 4.2.** By means of Marton's coupling argument [13], one can expect to prove the next Proposition 4.3. We are interested in transportation costs from $\mathcal{X}_1$ to $\mathcal{Y}_1$, from $\mathcal{X}_2$ to $\mathcal{Y}_2$ and from $\mathcal{X}_1\times\mathcal{X}_2$ to $\mathcal{Y}_1\times\mathcal{Y}_2$.
For any probability measure ν on the product space $\mathcal{Y}=\mathcal{Y}_1\times\mathcal{Y}_2$, let us write the desintegration of ν (conditional expectation) as follows: $\nu(dy_1dy_2)=\nu_1(dy_1)\nu_2^{y_1}(dy_2)$.

Proposition 4.3. *For all $\nu\in\mathcal{P}(\mathcal{Y}_1\times\mathcal{Y}_2)$,*

$$\mathcal{T}_{c_1\oplus c_2}(\mu_1\otimes\mu_2,\nu)\leq\mathcal{T}_{c_1}(\mu_1,\nu_1)+\int_{\mathcal{Y}_1}\mathcal{T}_{c_2}(\mu_2,\nu_2^{y_1})\,\nu_1(dy_1).\tag{4.4}$$

Recall that the relative entropy satisfies for all $\nu\in\mathcal{P}(\mathcal{Y}_1\times\mathcal{Y}_2)$,

$$H(\nu\mid\mu_1\otimes\mu_2)=H(\nu_1\mid\mu_1)+\int_{Y_1}H(\nu_2^{y_1}\mid\mu_2)\,\nu_1(dy_1)\tag{4.5}$$

which looks like (4.4).

Admitting Proposition 4.3 for a while, one can easily derive Theorem 4.2 as follows. Take $\mathcal{Y}_1 = \mathcal{X}_1$ and $\mathcal{Y}_2 = \mathcal{X}_2$. For all $\nu \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2)$,

$$\begin{aligned}
& \alpha_1 \square \alpha_2 (\mathcal{T}_{c_1 \oplus c_2}(\mu_1 \otimes \mu_2, \nu)) \\
\overset{(a)}{\leq} \; & \alpha_1 \square \alpha_2 \left(\mathcal{T}_{c_1}(\mu_1, \nu_1) + \int_{\mathcal{Y}_1} \mathcal{T}_{c_2}(\mu_2, \nu_2^{y_1}) \, \nu_1(dy_1) \right) \\
\overset{(b)}{\leq} \; & \alpha_1(\mathcal{T}_{c_1}(\mu_1, \nu_1)) + \alpha_2 \left(\int_{\mathcal{Y}_1} \mathcal{T}_{c_2}(\mu_2, \nu_2^{y_1}) \, \nu_1(dy_1) \right) \\
\overset{(c)}{\leq} \; & \alpha_1(\mathcal{T}_{c_1}(\mu_1, \nu_1)) + \int_{\mathcal{Y}_1} \alpha_2 \big(\mathcal{T}_{c_2}(\mu_2, \nu_2^{y_1}) \big) \, \nu_1(dy_1) \\
\overset{(d)}{\leq} \; & H(\nu_1 \mid \mu_1) + \int_{\mathcal{Y}_1} H(\nu_2^{y_1} \mid \mu_2) \, \nu_1(dy_1) \\
= \; & H(\nu \mid \mu_1 \otimes \mu_2).
\end{aligned}$$

Inequality (a) holds thanks to Proposition 4.3 since $\alpha_1 \square \alpha_2$ is increasing, (b) follows from the very definition of the inf-convolution, (c) follows from Jensen's inequality since α_2 is convex, (d) follows from the assumptions $\alpha_1(\mathcal{T}_1(\nu_1)) \leq H(\nu_1 \mid \mu_1)$ for all ν_1 and $\alpha_2(\mathcal{T}_2(\nu_2)) \leq H(\nu_2 \mid \mu_2)$ for all ν_2 (with obvious notations) and the last equality is (4.5). To complete the proof of Theorem 4.2, it remains to prove Proposition 4.3. This won't be achieved completely: a difficult measurability statement will only be conjectured.

Incomplete proof of Proposition 4.3. One first faces a nightmare of notations. It might be helpful to introduce random variables and see $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) = \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}_1 \times \mathcal{Y}_2)$ as the law of (X_1, X_2, Y_1, Y_2). One denotes $\pi_1 = \mathcal{L}(X_1, Y_1)$, $\pi_2^{x_1,y_1} \mathcal{L}(X_2, Y_2 \mid X_1 = x_1, Y_1 = y_1)$, $\pi_{X_2}^{x_1,y_1} = \mathcal{L}(X_2 \mid X_1 = x_1, Y_1 = y_1)$, $\pi_{Y_2}^{x_1,y_1} = \mathcal{L}(Y_2 \mid X_1 = x_1, Y_1 = y_1)$ $\pi_X = \mathcal{L}(X_1, X_2)$, $\pi_Y = \mathcal{L}(Y_1, Y_2)$ and so on.
Let us denote $P(\mu, \nu)$ the set of all $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ such that $\pi_X = \mu$ and $\pi_Y = \nu$, $P_1(\mu_1, \nu_1)$ the set of all $\eta \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{Y}_1)$ such that $\eta_{X_1} = \mu_1$ and $\eta_{Y_1} = \nu_1$ and $P_2(\mu_2, \nu_2)$ the set of all $\eta \in \mathcal{P}(\mathcal{X}_2 \times \mathcal{Y}_2)$ such that $\eta_{X_2} = \mu_2$ and $\eta_{Y_2} = \nu_2$.
We only consider couplings π such that under the law π

- $\mathcal{L}(X_1, X_2) = \mu$,
- $\mathcal{L}(Y_1, Y_2) = \nu$,
- Y_1 and X_2 are independent conditionally on X_1 and
- X_1 and Y_2 are independent conditionally on Y_1.

Optimizing over this collection of couplings leads us to

$$\mathcal{T}_c(\mu, \nu) \leq \inf_{\pi_1, \pi_2^\diamond} \int c_1 \oplus c_2(x_1, y_1, x_2, y_2) \, \pi_1(dx_1 dy_1) \pi_2^{x_1,y_1}(dx_2 dy_2)$$

where the infimum is taken over all $\pi_1 \in P_1(\mu_1, \nu_1)$ and all $\pi_2^\diamond = (\pi_2^{x_1,y_1}; x_1 \in \mathcal{X}_1, y_1 \in \mathcal{Y}_1)$ such that $\pi_2^{x_1,y_1} \in P_2(\mu_{X_2}^{x_1}, \nu_{Y_2}^{y_1})$ for π_1-almost every (x_1, y_1). As μ is a tensor product: $\mu = \mu_1 \otimes \mu_2$, we have $\mu_{X_2}^{x_1} = \mu_2$, π_1-a.e. so that $\pi_2^{x_1,y_1} \in P_2(\mu_2, \nu_{Y_2}^{y_1})$ for π_1-almost every (x_1, y_1).

Not being careful, one may write

$$\begin{aligned}
&\mathcal{T}_c(\mu,\nu)\\
\le\ & \inf_{\pi_1,\pi_2^\diamond} \int c_1\oplus c_2(x_1,y_1,x_2,y_2)\,\pi_1(dx_1dy_1)\pi_2^{x_1,y_1}(dx_2dy_2)\\
=\ & \inf_{\pi_1}\left[\int_{\mathcal{X}_1\times\mathcal{Y}_1} c_1\,d\pi_1 + \int_{\mathcal{X}_1\times\mathcal{Y}_1}\left(\inf_{\pi_2^\diamond}\int_{\mathcal{X}_2\times\mathcal{Y}_2} c_2(x_2,y_2)\pi_2^{x_1,y_1}(dx_2dy_2)\right)\pi_1(dx_1dy_1)\right]\\
\overset{(a)}{=}\ & \inf_{\pi_1}\left[\int_{\mathcal{X}_1\times\mathcal{Y}_1} c_1\,d\pi_1 + \int_{\mathcal{X}_1\times\mathcal{Y}_1}\left(\int_{\mathcal{X}_2\times\mathcal{Y}_2} c_2\,d\widehat{\pi}_2^{x_1,y_1}\right)\pi_1(dx_1dy_1)\right]\\
=\ & \inf_{\pi_1}\left[\int_{\mathcal{X}_1\times\mathcal{Y}_1} c_1\,d\pi_1 + \int_{\mathcal{X}_1\times\mathcal{Y}_1}\mathcal{T}_{c_2}\big(\mu_2,\nu_{Y_2}^{y_1}\big)\,\pi_1(dx_1dy_1)\right]\\
=\ & \inf_{\pi_1}\left\{\int_{\mathcal{X}_1\times\mathcal{Y}_1} c_1\,d\pi_1\right\} + \int_{\mathcal{Y}_1}\mathcal{T}_{c_2}(\mu_2,\nu_{Y_2}^{y_1})\,\nu_1(dy_1)\\
=\ & \mathcal{T}_{c_1}(\mu_1,\nu_1) + \int_{\mathcal{Y}_1}\mathcal{T}_{c_2}(\mu_2,\nu_{Y_2}^{y_1})\,\nu_1(dy_1)
\end{aligned}$$

which is the desired result.
On the right-hand side of equality (a), $\widehat{\pi}_2^{x_1,y_1}$ is a minimizer of $\pi_2^{x_1,y_1}\mapsto\int_{\mathcal{X}_2\times\mathcal{Y}_2} c_2\,d\pi_2^{x_1,y_1}$ subject to the constraint $\pi_2^{x_1,y_1}\in P_2(\mu_2,\nu_{Y_2}^{y_1})$. The general theory of optimal transportation insures that such a minimizer exists for each (x_1,y_1). And it might seem that the work is done.
But this is not true since one still has to prove that there exists a *measurable* mapping $(x_1,y_1)\mapsto\widehat{\pi}_2^{x_1,y_1}$. We now face a difficult problem that may possibly be solved by means of a measurable selection theorem, taking advantage of the pleasant property of tightness of any probability measure on a Polish space.
We withdraw this promising direct approach. □

4.3. **A complete indirect proof of Theorem 4.2.** It is based upon an indirect dual approach, making use of the characterization of Corollary 3.14 and follows the line of proof of ([10], Proposition 1.19).

Proof of Theorem 4.2. Recall that, provided that c is continuous nonnegative and satisfy (3.9), $Q^c\varphi(x)=\inf_{y\in\mathcal{X}}\{\varphi(y)+c(y,x)\}$ is in $B(\mathcal{X})$ whenever $\varphi\in B(\mathcal{X})$. We denote $Q_1=Q^{c_1}$, $Q_2=Q^{c_2}$ and $Q=Q^{c_1\oplus c_2}$.
By Corollary 3.14, the convex TCIs "$\alpha_1(\mathcal{T}_1)\le H_1$" and "$\alpha_2(\mathcal{T}_2)\le H_2$" which are supposed to hold are equivalent to

$$\int_{\mathcal{X}_1} e^{sQ_1\theta_1}\,d\mu_1 = \exp(\alpha_1^\circledast(s)+s\langle\theta_1,\mu_1\rangle),\quad \forall s\ge0,\forall\theta_1\in B(\mathcal{X}_1) \tag{4.6}$$

$$\int_{\mathcal{X}_2} e^{sQ_2\theta_2}\,d\mu_2 = \exp(\alpha_2^\circledast(s)+s\langle\theta_2,\mu_2\rangle),\quad \forall s\ge0,\forall\theta_2\in B(\mathcal{X}_2) \tag{4.7}$$

As by Lemma 4.1 $(\alpha_1\Box\alpha_2)^\circledast=\alpha_1^\circledast+\alpha_2^\circledast$, thanks to Corollary 3.14 again, all we have to prove is

$$\int_{\mathcal{X}_1\times\mathcal{X}_2} e^{sQ\varphi}\,d(\mu_1\otimes\mu_2)=\exp(\alpha_1^\circledast+\alpha_2^\circledast(s)+s\langle\varphi,\mu_1\otimes\mu_2\rangle),\quad \forall s\ge0,\forall\varphi\in C_b(\mathcal{X}_1\times\mathcal{X}_2) \tag{4.8}$$

Let us take $\varphi \in C_b(\mathcal{X}_1 \times \mathcal{X}_2)$. For all $(x_1, x_2) \in \mathcal{X}_1 \times \mathcal{X}_2$,

$$\begin{aligned} Q\varphi(x_1, x_2) &= \inf_{y_1 \in \mathcal{X}_1, y_2 \in Y_2} \{\varphi(y_1, y_2) + c_1(y_1, x_1) + c_2(y_2, x_2)\} \\ &= \inf_{y_1 \in \mathcal{X}_1} \left\{ \inf_{y_2 \in Y_2} \{\varphi(y_1, y_2) + c_2(y_2, x_2)\} + c_1(y_1, x_1) \right\} \\ &= \inf_{y_1 \in \mathcal{X}_1} \{\theta_{x_2}(y_1) + c_1(y_1, x_1)\} \\ &= Q_1 \theta_{x_2}(x_1) \end{aligned}$$

where

$$\theta_{x_2}(y_1) = Q_2 \varphi_{y_1}(x_2) = \inf_{y_2 \in Y_2} \{\varphi(y_1, y_2) + c_2(y_2, x_2)\} \tag{4.9}$$

with $\varphi_{y_1}(y_2) := \varphi(y_1, y_2)$. Hence, for all $s \geq 0$,

$$\begin{aligned} \int_{\mathcal{X}_1 \times \mathcal{X}_2} e^{sQ\varphi}\, d(\mu_1 \otimes \mu_2) &\overset{(a)}{=} \int_{\mathcal{X}_2} \left(\int_{\mathcal{X}_1} e^{sQ_1 \theta_{x_2}(x_1)}\, \mu_1(dx_1) \right) \mu_2(dx_2) \\ &\overset{(b)}{\leq} \int_{\mathcal{X}_2} e^{\alpha_1^{\circledast}(s) + s\langle \theta_{x_2}, \mu_1 \rangle}\, \mu_2(dx_2) \\ &\overset{(c)}{=} e^{\alpha_1^{\circledast}(s)} \int_{\mathcal{X}_2} \exp\left(s \int_{\mathcal{X}_1} Q_2 \varphi_{y_1}(x_2)\, \mu_1(dy_1) \right) \mu_2(dx_2) \end{aligned}$$

Equality (a) is justified since φ being bounded, $(x_1, x_2) \mapsto Q\varphi(x_1, x_2) = Q_1\theta_{x_2}(x_1)$ is jointly measurable.
Let us now prove the inequality (b). As φ and c are continuous, $(x_2, y_1) \mapsto \theta_{x_2}(y_1)$ is jointly upper semicontinuous as the infimum of a collection of continuous functions. Since $\theta_{x_2}(y_1) = Q_2\varphi_{y_1}(x_2)$ by (4.9), we have $\sup_{x_1, x_2} |\theta_{x_2}(y_1)| \leq \sup_{y_1} \sup |\varphi_{y_1}| = \sup |\varphi| < \infty$. Therefore, $(x_2, y_1) \mapsto \theta_{x_2}(y_1)$ is an upper semicontinuous bounded function. Consequently, one is allowed to invoke (4.6) to obtain $\int_{\mathcal{X}_1} e^{sQ_1\theta_{x_2}(x_1)}\, \mu_1(dx_1) \leq e^{\alpha_1^{\circledast}(s) + s\langle \theta_{x_2}, \mu_1 \rangle}$ for all x_2. Also note that $x_2 \mapsto \langle \theta_{x_2}, \mu_1 \rangle$ is measurable since $(x_2, y_1) \mapsto \theta_{x_2}(y_1)$ is jointly measurable and bounded.
The last equality (c) is simply (4.9).

Remark 4.10. If c_2 is only assumed to be lower semicontinuous, the joint measurability of $(x_2, y_1) \mapsto \theta_{x_2}(y_1)$ which has been used to prove inequality (b) is far from being clear. This is the reason why the cost functions are supposed to be continuous.

But for all x_2,

$$\begin{aligned} \int_{\mathcal{X}_1} Q_2 \varphi_{y_1}(x_2)\, \mu_1(dy_1) &= \int_{\mathcal{X}_1} \inf_{y_2 \in Y_2} \{\varphi(y_1, y_2) + c_2(y_2, x_2)\}\, \mu_1(dy_1) \\ &\leq \inf_{y_2 \in Y_2} \left\{ \int_{\mathcal{X}_1} \varphi(y_1, y_2)\, \mu_1(dy_1) + c_2(y_2, x_2) \right\} \\ &= Q_2 \overline{\varphi}(x_2) \end{aligned}$$

where $y_2 \mapsto \overline{\varphi}(y_2) = \int_{\mathcal{X}_1} \varphi(y_1, y_2)\, \mu_1(dy_1)$ is a continuous bounded function. Gathering our partial results leads us, for all $s \geq 0$, to the inequality (a) below

$$\begin{aligned}\int_{\mathcal{X}_1\times\mathcal{X}_2} e^{sQ\varphi}\, d(\mu_1 \otimes \mu_2) &\overset{(a)}{\leq} e^{\alpha_1^{\circledast}(s)} \int_{\mathcal{X}_2} e^{sQ_2\overline{\varphi}}\, d\mu_2 \\ &\overset{(b)}{\leq} e^{\alpha_1^{\circledast}(s)} e^{\alpha_2^{\circledast}(s)+s\langle \overline{\varphi},\mu_2\rangle} \\ &= e^{\alpha_1^{\circledast}(s)+\alpha_2^{\circledast}(s)+s\langle \varphi,\mu_1\otimes\mu_2\rangle}\end{aligned}$$

Inequality (b) is a consequence of (4.7). This is (4.8) and concludes the proof of the theorem. $\square$

4.4. **Product of n spaces.** The extension of Theorem 4.2 to the product of n spaces is as follows. Let $\mathcal{X}_1, \dots, \mathcal{X}_n$ be n Polish spaces and $\mu_1, \dots, \mu_n$ be probability measures on each of these spaces. On each space $\mathcal{X}_i$ let c_i be a cost function. The cost function on the product space $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$ is

$$c_1 \oplus \cdots \oplus c_n\big((x_1, \dots, x_n), (y_1, \dots, y_n)\big) = c_1(x_1, y_1) + \cdots + c_n(x_n, y_n)$$

Corollary 4.11. *Let us assume that the cost functions c_i are nonnegative continuous and satisfy (3.9). Suppose that the convex transportation cost inequalities*

$$\alpha_i(\mathcal{T}_{c_i}(\mu_i, \nu_i)) \leq H(\nu_i \mid \mu_i),\ \forall \nu_i \in \mathcal{P}(\mathcal{X}_i), \quad i = 1, \dots, n$$

hold with $\alpha_1, \dots, \alpha_n \in \mathcal{C}$. Then, on the product space $\mathcal{X}_1 \times \cdots \times \mathcal{X}_n$, we have the convex transportation cost inequality

$$\alpha_1 \square \cdots \square \alpha_n\big(\mathcal{T}_{c_1\oplus\cdots\oplus c_n}(\mu_1 \otimes \cdots \otimes \mu_n, \nu)\big) \leq H(\nu \mid \mu_1 \otimes \cdots \otimes \mu_n),\ \forall \nu \in \mathcal{P}(\mathcal{X}_1 \times \cdots \times \mathcal{X}_n)$$

where

$$\alpha_1 \square \cdots \square \alpha_n(t) = \inf\{\alpha_1(t_1) + \cdots + \alpha_n(t_n); t_1, \dots, t_n \geq 0 : t_1 + \cdots + t_n = t\},\ t \geq 0$$

is the inf-convolution of $\alpha_1, \dots, \alpha_n$.

Proof. It is a direct consequence of Theorem 4.2 which is proved by induction, noting that $\alpha_1 \square \cdots \square \alpha_n = (\alpha_1 \square \cdots \square \alpha_{n-1}) \square \alpha_n$ for all n. $\square$

In the special situation where the n TCIs are copies of a unique TCI on a Polish space $\mathcal{X}$ we have the following important result.

Theorem 4.12. *Let us assume that the cost function c is nonnegative continuous and satisfy (3.9). Suppose that the convex transportation cost inequality*

$$\alpha(\mathcal{T}_c(\mu, \nu)) \leq H(\nu \mid \mu),\ \forall \nu \in \mathcal{P}(\mathcal{X})$$

holds with $\alpha \in \mathcal{C}$. Then, on the product space $\mathcal{X}^n$, we have the following convex transportation cost inequality

$$n\alpha\left(\frac{\mathcal{T}_{c^{\oplus n}}(\mu^{\otimes n}, \zeta)}{n}\right) \leq H(\zeta \mid \mu^{\otimes n}),\ \forall \zeta \in \mathcal{P}(\mathcal{X}^n)$$

where $c^{\oplus n}\big((x_1, \dots, x_n), (y_1, \dots, y_n)\big) = c(x_1, y_1) + \cdots + c(x_n, y_n)$.

Proof. This is a direct application of Corollary 4.11, noting that $\alpha^{\square n}(t) = n\alpha(t/n)$. $\square$

About dimension-free tensorized convex TCIs. Let us say that a convex transportation cost inequality

$$\alpha\left(\mathcal{T}_c(\mu,\nu)\right) \le H(\nu \mid \mu), \quad \forall \nu \in \mathcal{P}(\mathcal{X}) \tag{4.13}$$

has the dimension-free tensorization property, if the inequality

$$\alpha\left(\mathcal{T}_{c^{\oplus n}}(\mu^{\otimes n},\zeta)\right) \le H(\zeta \mid \mu^{\otimes n}), \quad \forall \zeta \in \mathcal{P}(\mathcal{X}^n)$$

holds for all $n \in \mathbb{N}^*$.

Clearly, according to Theorem 4.12, if $\alpha \in \mathcal{C}$ is of the form $\alpha(t) = at$ with $a \ge 0$, then (4.13) has the dimension-free tensorization property.

Remark 4.14. Thanks to the same theorem, a seemingly weaker sufficient condition on α for (4.13) to be dimension-free is $\alpha(t) \le \inf_{n\ge1} n\alpha(t/n)$, $t \ge 0$. As α is in $\mathcal{C}$, $\alpha(t)/t$ is an increasing function so that $\alpha'(0) := \lim_{t\downarrow 0} \alpha(t)/t$ exists. It follows that $\lim_{n\to\infty} n\alpha(t/n) = \alpha'(0)t$ for all $t \ge 0$. Therefore, the condition $\alpha(t) \le \inf_{n\ge1} n\alpha(t/n)$, $t \ge 0$ is equivalent to $\alpha(t) \le \alpha'(0)t$, $t \ge 0$. But since α is convex, the converse inequality also holds, that is $\alpha(t) \ge \alpha'(0)t$, $t \ge 0$. Consequently α is of the form $\alpha(t) = at$ with $a \ge 0$.

Dimension free tensorization is a phenomenon that can only happen when dealing with *non-metric* cost functions. Indeed, we show in the following proposition, that convex T_1-inequalities having this property are all trivial.

Proposition 4.15. *Let $(\mathcal{X}, d)$ be a Polish space and $\mu \in \mathcal{P}(\mathcal{X})$. The convex transportation cost inequality*

$$\alpha\left(\mathcal{T}_d(\mu,\nu)\right) \le H(\nu \mid \mu), \quad \forall \nu \in \mathcal{P}(\mathcal{X}), \tag{4.16}$$

with $\alpha \in \mathcal{C}$ has the dimension free tensorization property if, and only if $\alpha = 0$ or μ is a Dirac mass.

Proof. If $\alpha = 0$, it is clear that (4.16) has the dimension free tensorization property. If μ is a Dirac mass, it is easy to see that (4.16) holds for every $\alpha \in \mathcal{C}$. Noting that a tensor product of Dirac measures is again a Dirac measure, the dimension-free tensorization property is established in this special case.
Now, suppose that (4.16) has the dimension-free tensorization property, with $\alpha \ne 0$ and let us prove that μ is a Dirac mass. According to Theorem 3.17, the following inequality

$$\log \int_{\mathcal{X}^n} e^{s(\varphi(x_1)+\cdots+\varphi(x_n)-n\langle\varphi,\mu\rangle)}\,\mu^{\otimes n}(dx_1\dots dx_n) \le \alpha^{\circledast}(s), \quad \forall s \ge 0$$

holds for all bounded 1-Lipschitz φ and all $n \ge 1$. As a consequence, denoting by Λ_φ the Log-Laplace of $\varphi(X) - \langle\varphi,\mu\rangle$, X of law μ, one has $\Lambda_\varphi \le \frac{1}{n}\alpha^{\circledast}$, for all $n \ge 1$, and so $\Lambda_\varphi \le 0$ on $\operatorname{dom}\alpha^{\circledast}$ (the effective domain of $\alpha^{\circledast}$). But by Jensen inequality, one obtains immediately $\Lambda_\varphi \ge 0$. Thus $\Lambda_\varphi \equiv 0$ on $\operatorname{dom}\alpha^{\circledast}$. As $\alpha \ne 0$, $[0,a[\subset \operatorname{dom}\alpha^{\circledast}$, for some $a > 0$. Considering $-\varphi$ instead of φ in the above reasoning yields that $\Lambda_\varphi \equiv 0$ on $]-a,a[$. This easily implies that μ_φ (the image of μ under the application φ) is a Dirac mass. Now, let us take a point x_0 in the support of μ and consider the bounded 1-Lipschitz function $\varphi_0(x) = d(x,x_0)\wedge 1$, $x \in \mathcal{X}$. As x_0 is in the support of μ, $\mu_{\varphi_0}([0,\varepsilon[) = \mu(\varphi_0 < \varepsilon) > 0$ for all $\varepsilon > 0$. As μ_{φ_0} is a Dirac mass, one thus has $\mu(\varphi_0 < \varepsilon) = 1$ for all $\varepsilon > 0$. This easily implies that $\mu = \delta_{x_0}$. $\square$

5. Integral criteria

Our aim in this section is to give integral criteria for a convex $\mathcal{T}$-inequality to hold.
Let us first note that when two $\mathcal{T}$-inequalities $\alpha_0(\mathcal{T}(\nu)) \le H(\nu \mid \mu)$, $\forall \nu \in \mathcal{N}$ and $\alpha_1(\mathcal{T}(\nu)) \le H(\nu \mid \mu)$, $\forall \nu \in \mathcal{N}$ hold, then we have the resulting new inequality $\alpha(\mathcal{T}(\nu)) \le H(\nu \mid \mu)$, $\forall \nu \in \mathcal{N}$ with

$$\alpha = \max(\alpha_0, \alpha_1). \tag{5.1}$$

This allows us to separate our investigation into two parts: obtaining α_0 and α_1 which control respectively the small (neighbourhood of $t = 0$) and large values of t (the other ones). Let us go on with some vocabulary.

5.1. **Transportation functions and deviation functions.** We introduce the following definitions. Recall that $\mathcal{T}$ is defined at (2.4).

Definition 5.2 (Transportation function)**.** A left continuous increasing function $\alpha : [0,\infty) \to [0,\infty]$ is called a *transportation function* for $\mathcal{T}$ in $\mathcal{N}$ if

$$\alpha(\mathcal{T}(\nu)) \le H(\nu \mid \mu), \quad \forall \nu \in \mathcal{N}.$$

This means that the $\mathcal{T}$-inequality (2.3) holds with α.

Definition 5.3 (Deviation function)**.** A left continuous increasing function $\alpha : [0,\infty) \to [0,\infty]$ is called a *deviation function* for $\mathcal{T}$ if

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\mathcal{T}(L_n) \ge t) \le -\alpha(t), \quad \forall t \ge 0.$$

These functions will be shortly called later transportation and deviation functions, without any reference to $\mathcal{T}$ and $\mathcal{N}$.

Remark 5.4. For $\mathcal{T}(L_n)$ to be measurable, it is assumed that Φ is a set of couples of *continuous* functions. Indeed,

$$\left\{ \mathbf{x} \in \mathcal{X}^n; \mathcal{T}\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\right) \le t \right\} = \bigcap_{\phi\in\Phi} \left\{ \mathbf{x} \in \mathcal{X}^n; \frac{1}{n}\sum_{i=1}^n \varphi(x_i) + \langle \psi, \mu\rangle \le t \right\}$$

is a closed set.

Note that an increasing function is left continuous if and only if it is lower semicontinuous. Clearly, the best transportation function is the left continuous version of the increasing function

$$t \mapsto \inf\{H(\nu \mid \mu); \nu \in \mathcal{N}, \mathcal{T}(\nu) \ge t\}, t \ge 0.$$

Similarly, the best deviation function is the left continuous version of the increasing function

$$t \mapsto -\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\mathcal{T}(L_n) \ge t) \in [0,\infty], t \ge 0.$$

Proposition 5.5. *Under the assumptions of Theorem 3.7, any deviation function α in the class $\mathcal{C}$ is a transportation function.*

Proof. Let $\alpha \in \mathcal{C}$ be a deviation function. Since $\mathcal{T}(L_n) \geq T_n^\phi$ for all $\phi \in \Phi$, we clearly have $\mathbb{P}(\mathcal{T}(L_n) \geq t) \geq \mathbb{P}(T_n^\phi \geq t)$ for all $t \geq 0$ and n. Therefore, for all ϕ, n and t, $\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T_n^\phi \geq t) \leq \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\mathcal{T}(L_n) \geq t) \leq -\alpha(t)$. This implies the statement (d) of Theorem 3.7, which in turn is equivalent to the statement (a) of Theorem 3.7, which is the desired result. □

5.2. **Controlling the large values of t.** In this subsection, it is assumed that the deviation and transportation functions are in $\mathcal{C}$.

Proposition 5.6. *The first statement is concerned with convex TCIs and the second one with convex $\mathcal{T}$-inequalities.*

(a) *If $\beta \in \mathcal{C}$ satisfies $\int_\mathcal{X} \exp[\beta(\int_\mathcal{X} c(x,y)\,\mu(dy))]\,\mu(dx) \leq A < \infty$ then*

$$\alpha(t) = \max(0, \beta(t) - \log A)), \quad t \geq 0$$

is a transportation function.

(b) *Let us suppose that α is a transportation function, then for all $(\psi, \varphi) \in \Phi$*

$$\int_\mathcal{X} \exp\left[\delta\alpha\left(\varphi(x) + \langle\psi, \mu\rangle\right)\right]\,\mu(dx) \leq \frac{1+\delta}{1-\delta} < \infty, \quad \forall 0 \leq \delta < 1.$$

Remarks.

- In (a), because of Jensen's inequality, one can take $A \geq \int_{\mathcal{X}^2} \exp\beta(c(x,y))\,\mu(dx)\mu(dy)$
- About (a), if $c = d \leq D < \infty$ is a lower semicontinuous bounded metric, one recovers that $\alpha(t) = \begin{cases} 0, & \text{if } t \leq D \\ +\infty, & \text{if } t > D \end{cases}$ is a transportation function, which is obvious.
- About (b) in the case of a TCI, let us note that $\sup_{(\psi,\varphi)\in\Phi_c}(\varphi(x) + \langle\psi,\mu\rangle) \leq \int_\mathcal{X} \sup_\phi(\varphi(x) + \psi(y))\,\mu(dy) \leq \int_\mathcal{X} c(x,y)\,\mu(dy)$ for all x. It follows that $\int_\mathcal{X} \exp\left[\delta\alpha\left((\varphi(x) + \langle\psi,\mu\rangle)\right)\right]\,\mu(dx) \leq \int_\mathcal{X} \exp\left[\delta\alpha\left(\int_\mathcal{X} c(x,y)\,\mu(dy)\right)\right]\,\mu(dx)$ for all $(\psi,\varphi) \in \Phi$. It would be pleasant to obtain the finiteness of an integral in terms of c. In the case where $c(x,y) = d(x,y)^p$, this will be performed below at Corollary 5.14.

Proof. Let us prove (a). As the product measure $\mu(dx)L_n(dy)$ has the right marginal measures, we get: $\mathcal{T}_c(\mu, L_n) := T_n \leq \int_{\mathcal{X}^2} c(x,y)\mu(dx)L_n(dy) = \langle c_\mu, L_n\rangle$ with $c_\mu(y) := \int_\mathcal{X} c(x,y)\,\mu(dx)$. It follows that for all $t \geq 0$,

$$\begin{aligned}
\mathbb{P}(T_n \geq t) &\leq \mathbb{P}(\langle c_\mu, L_n\rangle \geq t) \\
&\overset{(a)}{=} \mathbb{P}(\beta(\langle c_\mu, L_n\rangle) \geq \beta(t)) \\
&\overset{(b)}{\leq} \mathbb{P}(\langle \beta\circ c_\mu, L_n\rangle \geq \beta(t)) \\
&\overset{(c)}{=} \mathbb{P}(e^{\sum_{i=1}^n \beta\circ c_\mu(X_i)} \geq e^{n\beta(t)}) \\
&\overset{(d)}{\leq} e^{-n\beta(t)}\mathbb{E}e^{\sum_{i=1}^n \beta\circ c_\mu(X_i)} \\
&\overset{(e)}{=} \left[e^{-\beta(t)}\mathbb{E}e^{\beta\circ c_\mu(X)}\right]^n
\end{aligned}$$

where equality (a) follows from the monotony of β, (b) from the convexity of β and Jensen's inequality, (c) from the monotony of the exponential, (d) from Markov's inequality and (e) from the fact that (X_i) is an iid sequence. Finally,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T_n \geq t) \leq -\beta(t) + \log \int_{\mathcal{X}} e^{\beta\circ c_\mu}\, d\mu,\ \forall t \geq 0$$

which with Proposition 5.5 leads to the desired result.

Let us prove (b). As $\alpha \in \mathcal{C}$ is a transportation function, by Theorem 3.7 (keeping the notations of Theorem 3.7) we have for all

$$\alpha(t) \leq \Lambda_\phi^*(t),\ \forall \phi \in \Phi, \forall t \geq 0.$$

By Lemma 5.7 below, as Λ_ϕ^* is the Cramér transform of $\varphi(X) + \langle \psi, \mu\rangle$ we get

$$\mathbb{E} \exp\left[\delta \Lambda_\phi^*(\varphi(X) + \langle \psi, \mu\rangle)\right] \leq \frac{1+\delta}{1-\delta}, \forall 0 \leq \delta < 1, \forall \phi$$

Extending α with $\alpha(t) = 0$ for all $t \leq 0$, we obtain $\alpha \leq \Lambda_\phi^*$ for all ϕ, $\alpha \leq \Lambda_\phi^*$. Consequently we obtain

$$\int_{\mathcal{X}} \exp\left[\delta\alpha(\varphi(x) + \langle \psi, \mu\rangle)\right] \mu(dx) \leq \frac{1+\delta}{1-\delta}, \forall 0 \leq \delta < 1, \forall \phi$$

As $e^{\delta\alpha}$ is increasing, the desired result follows by monotone convergence. □

During the above proof, the following lemma has been used.

Lemma 5.7. *Let Z be a real random variable such that $\mathbb{E}e^{\lambda_o|Z|} < \infty$ for some $\lambda_o > 0$. Let h be its Cramér transform. Then for all $0 \leq \delta < 1$, $\mathbb{E}\exp[\delta h(Z)] \leq (1+\delta)/(1-\delta)$.*

Proof. This result with the upper bound $2/(1-\delta)$ instead of $(1+\delta)/(1-\delta)$ can be found in ([6], Lemma 5.1.14). For a proof of the improvement with $(1+\delta)/(1-\delta)$ see [9]. □

Corollary 5.8. *In this statement d is a lower semicontinuous semimetric and c is a lower semicontinuous cost function such that $c(x,x) = 0$ for all $x \in \mathcal{X}$.*

(a) *Suppose that there exists a nonnegative measurable function χ such that*

$$c \leq \chi \oplus \chi.$$

Let $\gamma \in \mathcal{C}$ be such that $\int_{\mathcal{X}} \exp[\gamma \circ \chi(x)]\, \mu(dx) \leq B < \infty$, then for any $x_o \in \mathcal{X}$

$$t \mapsto 2\max(0, 2\gamma(t/4) - \gamma\circ\chi(x_o) - \log B), \quad t \geq 0$$

is a transportation function for c.

(b) *Suppose that there exists $\theta \in \mathcal{C}$ such that*

$$\theta(d) \leq c.$$

If $\alpha \in \mathcal{C}$ is a transportation function for c, then

$$\int_{\mathcal{X}} \exp[u\ \alpha\circ\theta(d(x_o,x)/2)]\, \mu(dx) < \infty$$

for all $x_o \in \mathcal{X}$ and all $0 \leq u < 2$.

Proof. We begin with the case where $c = d$, $\chi(x) = d(x_o, x)$ and $\theta(d) = d$.

The case $c = d$. To prove (a) with $\chi(x) = d(x_o, x)$, we apply statement (a) of Proposition 5.6. Let β be in the class $\mathcal{C}$. We have for all $x_o \in \mathcal{X}$

$$\begin{aligned}
\int_{\mathcal{X}} \exp\left[\beta\Big(\int_{\mathcal{X}} d(x,y)\,\mu(dy)\Big)\right]\mu(dx) &\le \int_{\mathcal{X}^2} \exp[\beta(d(x,y))]\,\mu(dx)\mu(dy)\\
&\le \int_{\mathcal{X}^2} \exp\left[\beta\Big(\frac{2d(x_o,x)+2d(x_o,y)}{2}\Big)\right]\mu(dx)\mu(dy)\\
&\le \int_{\mathcal{X}^2} \exp[\beta(2d(x_o,x))/2+\beta(2d(x_o,y))/2]\,\mu(dx)\mu(dy)\\
&= \left(\int_{\mathcal{X}} \exp\left[\frac{\beta(2d(x_o,x))}{2}\right]\mu(dx)\right)^2\\
&:= A
\end{aligned}$$

Taking, $\gamma(t) = \beta(2t)/2$, one gets $A = B^2$ and

$$t \mapsto \max(0, \beta(t) - \log A) = \max(0, 2\gamma(t/2) - 2\log B) \tag{5.9}$$

is a transportation function for $c = d$.

Now, let us prove (b). Thanks to Kantorovich-Rubinstein equality (2.2) one can take $\Phi = \{(-\varphi, \varphi); \|\varphi\|_{\mathrm{Lip}} \le 1, \varphi \text{ bounded}\}$. Because of Proposition 5.6-(b), we have for all bounded φ with $\|\varphi\|_{\mathrm{Lip}} \le 1$:

$$\int_{\mathcal{X}} \exp[\delta\alpha(\varphi(x) - \langle\varphi,\mu\rangle)]\,\mu(dx) \le (1+\delta)/(1-\delta), \forall 0 \le \delta < 1.$$

The function $\varphi(x) = d(x_o, x)$ is 1-Lipschitz but it is not bounded in general. Let us introduce an approximation procedure. For all $k \ge 0$, with $m := \int_{\mathcal{X}} d(x_o,y)\,\mu(dy)$, we have

$$\begin{aligned}
\int_{\mathcal{X}} \exp[\delta\alpha((d(x_o,x)\wedge k) - m]\,\mu(dx) &\le \int_{\mathcal{X}} \exp\left[\delta\alpha((d(x_o,x)\wedge k) - \int_{\mathcal{X}}[d(x_o,y)\wedge k]\,\mu(dy)\right]\mu(dx)\\
&\le (1+\delta)/(1-\delta).
\end{aligned}$$

By monotone convergence, one concludes that for all $0 \le \delta < 1$,

$$\int_{\mathcal{X}} \exp[\delta\alpha(d(x_o,x) - m]\,\mu(dx) \le (1+\delta)/(1-\delta).$$

As

$$\begin{aligned}
2\delta\alpha(d(x_o,x)/2) &= 2\delta\alpha\left(\frac{d(x_o,x)-m}{2} + \frac{m}{2}\right)\\
&\le \delta[\alpha(d(x_o,x)-m) + \alpha(m)],
\end{aligned}$$

one sees that

$$\int_{\mathcal{X}} \exp[2\delta\alpha(d(x_o,x)/2)]\,\mu(dx) \le e^{\delta\alpha(m)}\int_{\mathcal{X}} \exp[\delta\alpha(d(x_o,x)-m]\,\mu(dx) \le e^{\delta\alpha(m)}(1+\delta)/(1-\delta)$$

which leads to

$$\int_{\mathcal{X}} \exp[2\delta\alpha(d(x_o,x)/2)]\,\mu(dx) < \infty \tag{5.10}$$

The general case. Let us prove (a). It is clear that $c(x,y) \le d_\chi(x,y)$ where d_χ is the semimetric defined by

$$d_\chi(x,y) = \mathbf{1}_{x\neq y}(\chi(x)+\chi(y)). \tag{5.11}$$

Remark 5.12. If χ admits two or more zeros, d_χ is a semimetric. Otherwise it is a metric. In the often studied case where $c = d^p$ with d a metric and $p \ge 1$, one takes $\chi(x) = 2^{p-1}d(x_o,x)^p$ (see the proof of Corollary 5.14 below) and d_χ is a metric.

Of course, for all $\nu \in \mathcal{N} = \mathcal{P}_\chi = \{\nu \in \mathcal{P}(\mathcal{X}); \int_{\mathcal{X}} \chi(x)\,\nu(dx) < \infty\}$, we have

$$\mathcal{T}_c(\nu) \le \mathcal{T}_{d_\chi}(\nu).$$

Therefore, any transportation function for d_χ is a transportation function for c. This easy but powerful trick is borrowed from the monograph by C. Villani ([17], Proposition 7.10). It has been proved at (5.9) that if $\int_{\mathcal{X}} \exp[\beta(d_\chi(x_o,x)]\,\mu(dx) \le C < \infty$ for some function $\beta \in \mathcal{C}$, then $\max(0, 2\beta(t/2) - 2\log C)$ is a transportation function for d_χ.
Taking $\beta(t) = 2\gamma(t/2)$, with convexity we have

$$\beta(d_\chi(x_o,x)) \le \gamma\circ\chi(x_o) + \gamma\circ\chi(x) \tag{5.13}$$

so that $\int_{\mathcal{X}} \exp[\beta(d_\chi(x_o,x)]\,\mu(dx) \le e^{\gamma\circ\chi(x_o)}B = C$. This leads us to $\max(0, 2\beta(t/2) - 2\log C) = 2\max(0, 2\gamma(t/4) - \gamma\circ\chi(x_o) - \log B)$ which is the desired result.

Let us prove (b). Because of Jensen's inequality, it is easy to show that $\theta(\mathcal{T}_d) \le \mathcal{T}_c$. As α is a transportation function for c, it follows that $\alpha\circ\theta$ is a transportation function for $\mathcal{T}_d$. Applying the already proved result (5.10) with $\alpha\circ\theta$ instead of α completes the proof of the corollary. $\square$

Now, we consider an important special case of convex TCI.

Corollary 5.14 ($c = d^p$)**.** *In this statement $c = d^p$ where d is a lower semicontinuous metric and $p \ge 1$.*

(a) *Let $\gamma \in \mathcal{C}$ be such that $\int_{\mathcal{X}} \exp[\gamma(d^p(x_o,y))]\,\mu(dy) \le B < \infty$ for some $x_o \in \mathcal{X}$, then*

$$t \mapsto \max(0, 2\gamma(2^{-p}t) - 2\log B), \quad t \ge 0$$

is a transportation function.

(b) *If $\alpha \in \mathcal{C}$ is a transportation function, then*

$$\int_{\mathcal{X}} \exp[u\ \alpha(2^{-p}d^p(x_o,x))]\,\mu(dx) < \infty$$

for all $x_o \in \mathcal{X}$ and all $0 \le u < 2$.

Proof. This is Corollary 5.8 with $\chi(x) = 2^{p-1}d^p(x_o,x)$, $\theta(d) = d^p$ and the following improvement in the treatment of the inequality (5.13). One can write $\beta(d_\chi(x_o,x)) \le \gamma\circ\chi(x_o) + \gamma\circ\chi(x) = \gamma\circ\chi(x)$ since $\gamma\circ\chi(x_o) = 0$ in this situation. As a consequence $\max(0, 2\gamma(2^{-p}t) - 2\log B)$ is a transportation function, which is a little better than its counterpart in Corollary 5.8. $\square$

Remark 5.15. It is known that the standard Gaussian measure μ on $\mathbb{R}$ satisfies T_2 which is the TCI with $c(x,y) = (x-y)^2$ and the transportation function $\alpha(t) = t/2$ (see [16]). As a consequence of Corollary 5.14-b, for all $p > 2$, there is no function α in $\mathcal{C}$ except $\alpha \equiv 0$ which is a transportation function for the standard Gaussian measure and the cost function $|x-y|^p$.

5.3. **Controlling the small values of t.** We are going to prove a general result for the behaviour of a transportation function in the neighbourhood of zero. By a general result, it is meant that μ is not specified. As a consequence, it will only be shown that under the assumption that $c \le \chi \oplus \chi$ where $\int_{\mathcal{X}} e^{\delta_o \chi}\, d\mu < \infty$ for some $\delta_o > 0$, there are tranportation functions which are larger than some quadratic function around zero. Obtaining better results in this direction is difficult and requires more stringent restrictions on the reference probability measure μ.

Proposition 5.16. *Let c be a cost function satisfying (3.9) and $c \le \chi \oplus \chi$ for some nonnegative measurable function χ satisfying $\int_{\mathcal{X}} e^{\delta_o \chi}\, d\mu < \infty$ for some $\delta_o > 0$. Then, $\|\chi\|_\rho$ is finite and*

$$\alpha_o(t) = \left(\sqrt{t/\|\chi\|_\rho + 1} - 1\right)^2,\ t \ge 0$$

is a transportation function for c and μ.
In particular, for all $a \ge 0$ such that $\int_{\mathcal{X}} e^{a\chi}\, d\mu \le 2$, $t \mapsto (\sqrt{at+1} - 1)^2$ is a transportation function.

Note that $(\sqrt{at+1} - 1)^2 = a^2t^2/4 + o_{t\to 0}(t^2) = at - 2\sqrt{at} + 2 + o_{t\to\infty}(1)$.
The Orlicz norm $\|\chi\|_\rho$ is defined at (3.21).

Proof. Because of our assumptions, we have $\mathcal{T}_c \le \mathcal{T}_{d_\chi}$, see (5.11). Hence, it is enough to show that α_o is a transportation function for d_χ. But this follows from Lemma 5.17 below and Corollary 3.24.
The last statement follows from a simple manipulation on the definition of the Orlicz norm $\|\chi\|_\rho$. $\square$

The following lemma has been used in the previous proof.

Lemma 5.17. *For all μ and ν in $\mathcal{P}_\chi := \{\nu \in \mathcal{P}(\mathcal{X}); \int_{\mathcal{X}} \chi\, d\nu < \infty\}$, we have*

$$\mathcal{T}_{d_\chi}(\mu, \nu) = \|\chi \cdot (\mu - \nu)\|_{\mathrm{TV}}.$$

Proof. By Kantorovich-Rubinstein's equality (2.2), we have $\mathcal{T}_{d_\chi}(\mu,\nu) = \sup\{\int_{\mathcal{X}} \varphi\, d(\nu - \mu); \varphi \in B(\mathcal{X}), \|\varphi\|_{\mathrm{Lip}} \le 1\}$ where $\|\varphi\|_{\mathrm{Lip}} \le 1$ is equivalent to $|\varphi(x) - \varphi(y)| \le d_\chi(x, y)$ for all x, y. One can prove without trouble (see [9]) that this is equivalent to $|\varphi(x) - a| \le \chi(x), \forall x$ for some real a. Therefore,

$$\begin{aligned}
\mathcal{T}_{d_\chi}(\mu,\nu) &= \sup\left\{\int_{\mathcal{X}} \varphi\, d(\nu - \mu); \varphi \in B(\mathcal{X}) : |\varphi| \le \chi\right\} \\
&= \sup_{k \ge 1} \sup\left\{\int_{\mathcal{X}} (\chi \wedge k)\theta\, d(\nu - \mu); \theta \in B(\mathcal{X}) : |\theta| \le 1\right\} \\
&= \|\chi \cdot (\mu - \nu)\|_{\mathrm{TV}}
\end{aligned}$$

which is the desired result. $\square$

5.4. **An application: T_1-inequalities.** A T_1-inequality is a TCI with $c = d$. Let us denote $\mathcal{P}_d(\mathcal{X}) = \{\nu \in \mathcal{P}(\mathcal{X}); \int_{\mathcal{X}} d(x_*, x)\, \nu(dx) < \infty$ for some (and therefore all) $x_* \in \mathcal{X}\}$. Suppose that μ is in $\mathcal{P}_d(\mathcal{X})$. The function α is said to satisfy the T_1-inequality for d and μ if

$$\alpha(\mathcal{T}_d(\mu, \nu)) \le H(\nu \mid \mu),\ \forall \nu \in \mathcal{P}_d(\mathcal{X}). \tag{5.18}$$

Theorem 5.19 (T_1-inequalities)**.** *Let d be a lower semicontinuous metric. Suppose that $a \geq 0$ satisfies $\int_{\mathcal{X}} e^{ad(x_o,x)}\,\mu(dx) \leq 2$ for some $x_o \in \mathcal{X}$ and that $\gamma \in \mathcal{C}$ satisfies $\int_{\mathcal{X}} e^{\gamma(d(x_1,x))}\,\mu(dx) \leq B < \infty$ for some $x_1 \in \mathcal{X}$, then*

$$\alpha(t) = \max\Big((\sqrt{at+1}-1)^2, 2\gamma(t/2) - 2\log B\Big),\ t \geq 0$$

satisfies (5.18).
Conversely, if a function α in the class $\mathcal{C}$ satisfies (5.18), then

$$\int_{\mathcal{X}} \exp[u\ \alpha(d(x_*,x)/2)]\,\mu(dx) < \infty$$

for all $x_* \in \mathcal{X}$ and all $0 \leq u < 2$.

Proof. Gathering Corollary 5.14-a, Proposition 5.16 and the trick (5.1) gives us the first statement. The second statement is a particular instance of Corollary 5.14-b. ☐

Note that by Proposition 3.19 we know that it is impossible that α escapes from a quadratic growth at the origin.
Theorem 5.19 extends the integral criteria for the usual $T_1(C)$-inequality in [7] and [**?**]. Nevertheless, the control of the constant C is handled more carefully in these cited papers. In a forthcoming paper (see the PhD manuscript [9]), one of the author has obtained the following result which is very much in the spirit of [7] and [**?**].

Theorem 5.20. *Suppose that $c(x,y) = d^p(x,y)$, that α satisfies (3.20) for some $a > 0$ and that $\alpha^{\circledast}$ is unbounded on its effective domain. Then, the following statements are equivalent :*

- *There exists $b_1 > 0$ such that $\alpha\left(b_1 \mathcal{T}_{d^p}(\nu,\mu)\right) \leq H(\nu|\mu)$ for all $\nu \in \mathcal{P}(\mathcal{X})$ such that $\int_{\mathcal{X}} d^p(x_o,x)\,\mu(dx) < \infty$*
- *There exists $b_2 > 0$ such that $\iint_{\mathcal{X}^2} e^{\alpha(b_2 d^p(x,y))}\,\mu(dx)\mu(dy) < +\infty$.*

Further details concerning the relation between b_1 and b_2 can be found in [9].

6. Some applications: concentration of measure and deviations of empirical processes

In this section, we give some applications of T_1-inequalities. The first application, Theorem 6.3 is an easy extension of a well known result of K. Marton. The second one, Theorem 6.10 is more original and concerns the deviations of empirical processes.
In the whole section, d is a metric on $\mathcal{X}$ which turns $(\mathcal{X}, d)$ into a Polish space.

6.1. **A basic lemma.**

Theorem 6.3 and Theorem 6.10 both rely on the following elementary lemma.

Lemma 6.1. *Let $\mu \in \mathcal{P}(\mathcal{X})$ be such that $\int_{\mathcal{X}} d(x_o,x)\,\mu(dx) < +\infty$, for all $x_0 \in \mathcal{X}$, and suppose that the T_1- inequality*

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

holds. Then, for all 1-Lipschitz function φ, one has

$$\mu\left(\varphi \geq \langle\varphi,\mu\rangle + t\right) \leq e^{-\alpha(t)}, \quad \forall t > 0. \tag{6.2}$$

Proof. Let φ a 1-Lipschitz function. For every $n \geq 1$, let us consider $\varphi_n = \varphi \vee n \wedge -n$. According to point b. of Theorem 3.17, one has

$$\Lambda_{\varphi_n}(s) := \log \int_{\mathcal{X}} e^{s(\varphi_n - \langle \varphi_n, \mu \rangle)}\, d\mu \leq \alpha^{\circledast}(s), \quad \forall s \geq 0.$$

By dominating convergence, $\langle \varphi_n, \mu \rangle \xrightarrow[n \to +\infty]{} \langle \varphi, \mu \rangle$. Thus by Fatou's lemma, one has

$$\Lambda_{\varphi}(s) := \log \int_{\mathcal{X}} e^{s(\varphi - \langle \varphi, \mu \rangle)}\, d\mu \leq \alpha^{\circledast}(s), \quad \forall s \geq 0.$$

Now, thanks to Chebychev argument, one has for all $t \geq 0$:

$$\mu\left(\varphi \geq \langle \varphi, \mu \rangle + t\right) \leq \inf_{s \geq 0} \int_{\mathcal{X}} e^{s(\varphi - \langle \varphi, \mu \rangle - t)}\, d\mu \leq \inf_{s \geq 0} e^{\alpha^{\circledast}(s) - st} = e^{-\alpha(t)}.$$

□

6.2. **T_1-inequalities and concentration of measure.** Let us recall that for a given probability measure μ on a Polish space $\mathcal{X}$, the concentration function of μ is defined by

$$\theta_\mu(r) = \sup\{1 - \mu(A^r) : A \text{ borel set such that } \mu(A) \geq 1/2\}, \quad \forall r > 0,$$

where

$$A^r := \{x \in \mathcal{X} : d(x, A) \leq r\}.$$

One says that θ is a concentration function for μ, if there is $r_0 \geq 0$ such that

$$\theta_\mu(r) \leq \theta(r), \quad \forall r \geq r_0,$$

or equivalently

$$\mu(A^r) \geq 1 - \theta(r), \quad \forall r \geq r_0, \quad \forall A \text{ Borel set.}$$

Roughly speaking, the following theorem states that if α is a T_1-transportation function for μ then $e^{-\alpha}$ is a concentration function for μ. This link between transportation cost inequality and concentration inequality was first noticed by K. Marton, see [12]. Her result extends as follows.

Theorem 6.3. *Let $\mu \in \mathcal{P}(\mathcal{X})$ be such that $\int_{\mathcal{X}} d(x_o, x)\, \mu(dx) < +\infty$ for all $x_0 \in \mathcal{X}$, and suppose that the T_1-inequality*

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

holds with an unbounded $\alpha \in \mathcal{C}$. Then for all measurable A with $\mu(A) > 0$, one has the following concentration of measure inequality :

$$\mu(A^r) \geq 1 - e^{-\alpha(r - r_A)}, \quad \forall r \geq r_A, \tag{6.4}$$

where $r_A := \alpha^{-1}(-\log \mu(A))$.

The following proof is different from Marton's original argument. Our proof is based on deviation arguments while Marton's one is based on transportation. For a proof using Marton's concentration arguments see Proposition VI.81 in [9].

Proof. The function $x \mapsto d(x, A)$ is 1-Lipschitz. Thus, according to Lemma 6.1,

$$\mu(d(\cdot, A) \geq t + \langle d(\cdot, A), \mu\rangle) \leq e^{-\alpha(t)}, \quad \forall t \geq 0.$$

In order to derive (6.4), the only thing to do is to show that $\langle d(\cdot, A), \mu\rangle \leq \alpha^{-1}(-\log \mu(A))$. Let $\nu \in \mathcal{P}(\mathcal{X})$ be such that $\nu(A) = 1$. According to the T_1-inequality satisfied by μ, one has

$$\int_{\mathcal{X}} d(\cdot, A)\, d\mu = \int_{\mathcal{X}} d(\cdot, A)\, d\mu - \int_{\mathcal{X}} d(\cdot, A)\, d\nu \leq \mathcal{T}_d(\mu, \nu) \leq \alpha^{-1}(H(\nu \mid \mu)).$$

Thus,

$$\langle d(\cdot, A), \mu\rangle \leq \alpha^{-1} \left(\inf \left\{H(\nu \mid \mu) : \nu(A) = 1\right\}\right).$$

Let $\mu_A \in \mathcal{P}(\mathcal{X})$ be defined by $d\mu_A = \frac{\mathbf{1}_A}{\mu(A)} d\mu$; clearly $\mu_A(A) = 1$, so

$$\inf \left\{H(\nu \mid \mu) : \nu(A) = 1\right\} \leq H(\mu_A \mid \mu). \tag{6.5}$$

An easy computation yields $H(\mu_A \mid \mu) = -\log \mu(A)$. □

Note that $d(\cdot, A)$ is unbounded so that the inequality $\int_{\mathcal{X}} d(\cdot, A)\, d\mu - \int_{\mathcal{X}} d(\cdot, A)\, d\nu \leq \mathcal{T}_d(\mu, \nu)$ needs to be justified. Let π be a probability on $\mathcal{X}^2$ with marginals μ and ν, then $\int_{\mathcal{X}} d(\cdot, A)\, d\mu - \int_{\mathcal{X}} d(\cdot, A)\, d\nu = \iint_{\mathcal{X}^2} d(x, A) - d(y, A)\, \pi(dxdy) \leq \iint_{\mathcal{X}^2} d(x, y)\, \pi(dxdy)$. Optimizing in π leads to the desired result.

Some comments. In Marton's approach, the probability measure μ_A plays also a great role. Thanks to our approach, this role can be further explained. The choice of μ_A is optimal in the sense that (6.5) holds with equality:

$$\inf \left\{H(\nu \mid \mu) : \nu(A) = 1\right\} = H(\mu_A \mid \mu). \tag{6.6}$$

In other words, μ_A is Csiszár's I-projection of μ on $\{\nu \in \mathcal{P}(\mathcal{X}) : \nu(A) = 1\}$, see [4, 5]. If ν is such that $\nu(A) = 1$, one has

$$\begin{aligned} H(\nu \mid \mu) &= H(\nu \mid \mu_A) + \int_{\mathcal{X}} \log \frac{d\mu_A}{d\mu}\, d\nu \\ &= H(\nu \mid \mu_A) + \int_{\mathcal{X}} \log \mathbf{1}_A\, d\nu - \log \mu(A) \\ &= H(\nu \mid \mu_A) + H(\mu_A \mid \mu), \end{aligned}$$

where the last equality follows from $\int_{\mathcal{X}} \log \mathbf{1}_A\, d\nu = 0$ and $H(\mu_A \mid \mu) = -\log \mu(A)$. This proves (6.6).

6.3. **T_1-inequalities and deviations bounds for empirical processes.** Lemma 6.1 together with the tensorization property of Theorem 4.12 immediately implies the following

Lemma 6.7. *Let $\mu \in \mathcal{P}(\mathcal{X})$ be such that $\int_{\mathcal{X}} d(x_o, x)\, \mu(dx) < +\infty$, for all $x_0 \in \mathcal{X}$, and suppose that the T_1-inequality*

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

holds. Then for all function $Z : \mathcal{X}^n \to \mathbb{R}$ which is $1/n$-Lipschitz with respect to the metric $d^{\oplus n}$, one has

$$\mu^{\otimes n}\left(Z \geq \langle \mu, Z\rangle + t\right) \leq e^{-n\alpha(t)}, \quad \forall t \geq 0 \tag{6.8}$$

Let us consider a class $\mathcal{G}$ of 1-Lipschitz functions on $\mathcal{X}$, and X_i an iid sample of law μ. Let $Z_n^{\mathcal{G}}$ be defined by

$$Z_n^{\mathcal{G}} := \sup_{\varphi\in\mathcal{G}} \left\{ \left| \frac{1}{n}\sum_{i=1}^n \varphi(X_i) - \int_{\mathcal{X}} \varphi\, d\mu \right| \right\}. \tag{6.9}$$

As $0 \leq Z_n^{\mathcal{G}} = \sup_{\varphi\in\mathcal{G}} \left\{ \left| \int_{\mathcal{X}} \varphi\, dL_n - \int_{\mathcal{X}} \varphi\, d\mu \right| \right\} \leq \mathcal{T}_d(L_n, \mu)$, one has $Z_n^{\mathcal{G}} \in [0, +\infty[$. Further, as a supremum of $1/n$-Lipschitz functions, the function

$$(x_1, \ldots, x_n) \mapsto \sup_{\varphi\in\mathcal{G}} \left\{ \left| \frac{1}{n}\sum_{i=1}^n \varphi(x_i) - \int_{\mathcal{X}} \varphi\, d\mu \right| \right\}$$

is $1/n$-Lipschitz too. This implies in particular that $Z_n^{\mathcal{G}}$ is measurable. The random variable $Z_n^{\mathcal{G}}$ is called an empirical process. Applying Lemma 6.7, one immediately obtains the following theorem.

Theorem 6.10. *Let $\mu \in \mathcal{P}(\mathcal{X})$ be such that $\int_{\mathcal{X}} d(x_o, x)\, \mu(dx) < +\infty$, for all $x_0 \in \mathcal{X}$, and suppose that the T_1-inequality*

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

holds. If $\mathcal{G}$ is a class of 1-Lipschitz functions on $\mathcal{X}$ then the empirical process $Z_n^{\mathcal{G}}$ defined by (6.9) satisfies the following inequality

$$\mathbb{P}\left(Z_n^{\mathcal{G}} \geq \mathbb{E}\left[Z_n^{\mathcal{G}}\right] + t\right) \leq e^{-n\alpha(t)}, \quad \forall t \geq 0. \tag{6.11}$$

The literature about the deviations of empirical processes is huge. For a good overview of this subject, one can read P. Massart's Saint-Flour lecture notes [14].
Now, if $(\mathcal{X}, \|\cdot\|)$ is a Banach space, and $\mu \in \mathcal{P}(\mathcal{X})$ such that $\int_{\mathcal{X}} \|x\|\, d\mu < +\infty$ then taking $\mathcal{G} = \{\ell \in \mathcal{X}^* : \|\ell\|_{\mathcal{X}^*} = 1\}$, where $\mathcal{X}^*$ is the topological dual space of $\mathcal{X}$, one obtains

$$Z_n^{\mathcal{G}} = \left\| \frac{1}{n}\sum_{i=1}^n X_i - \int_{\mathcal{X}} x\, d\mu \right\|,$$

where $\int_{\mathcal{X}} x\, \mu(dx)$ is well defined in the Bochner sense. In this special case, we have the following result.

Theorem 6.12. *Let $\mu \in \mathcal{P}(\mathcal{X})$ be such that $\int_{\mathcal{X}} \|x\|\, \mu(dx) < +\infty$, and suppose that the T_1-inequality*

$$\alpha\left(\mathcal{T}_{\|\cdot\|}(\mu, \nu)\right) \leq H(\nu \mid \mu), \forall \nu \in \mathcal{P}(\mathcal{X}),$$

holds. If X_i is an iid sequence of law μ, then letting $Z_n = \left\| \frac{1}{n}\sum_{i=1}^n X_i - \int_{\mathcal{X}} x\, d\mu \right\|$, one has

$$\mathbb{P}\left(Z_n \geq \mathbb{E}\left[Z_n\right] + t\right) \leq e^{-n\alpha(t)}, \quad \forall t \geq 0. \tag{6.13}$$

Remark 6.14. In order to obtain precise deviations results for $Z_n^{\mathcal{G}}$ (resp. Z_n), one must be able to estimate the term $\mathbb{E}\left[Z_n^{\mathcal{G}}\right]$ (resp. $\mathbb{E}\left[Z_n\right]$).

Let us give some examples.

Example 1. Quantitative versions of Sanov theorem. Suppose that $\mathcal{G}$ is the set of all bounded 1-Lipschitz functions on $\mathcal{X}$, then $Z_n^{\mathcal{G}} = \mathcal{T}_d(L_n, \mu)$, see (2.2).
The following theorem is Theorem 10.2.1 of [15] (volume II).

Theorem 6.15. *Let μ be a probability measure on $\mathbb{R}^q$ (equipped with its usual euclidean norm $\|\cdot\|_2$) such that*

$$c := \int \|x\|_2^{q+5}\, d\mu < +\infty. \tag{6.16}$$

Then, there is $D > 0$ depending only on c and q, such that

$$\mathbb{E}\left[\mathcal{T}_{d_2}(L_n,\mu)\right] \leq D n^{-\frac{1}{q+4}}, \tag{6.17}$$

where d_2 is the metric associated to $\|\cdot\|_2$.

Thanks to this result, one obtains the following quantitative version of Sanov theorem :

Corollary 6.18. *Let μ be a probability on $\mathbb{R}^q$, satisfying (6.16) and the T_1-inequality*

$$\alpha\left(\mathcal{T}_{d_2}(\mu,\nu)\right) \leq H(\nu \mid \mu), \quad \forall \nu \in \mathcal{P}(\mathbb{R}^q),$$

where d_2 is the usual euclidean metric on $\mathbb{R}^q$. Then, the following inequality holds :

$$\mathbb{P}\left(\mathcal{T}_{d_2}(L_n,\mu) \geq t\right) \leq \exp\left(-n\alpha\left(t - \frac{D}{n^{\frac{1}{q+4}}}\right)\right), \quad \forall t > 0, \quad \forall n \geq \left(\frac{D}{t}\right)^{q+4},$$

where D is the constant of (6.17).

In [3], F. Bolley, A. Guillin and C. Villani have also obtained a quantitative version of Sanov theorem with alternative arguments.

Example 2. Deviations bounds for empirical means. Let $\mathcal{X}$ be a Banach space and consider

$$Z_n = \left\| \frac{1}{n}\sum_{i=1}^n X_i - \int_{\mathcal{X}} x\, d\mu \right\|, \tag{6.19}$$

where X_i is an iid sequence of law μ. In order to control the term $\mathbb{E}[Z_n]$, a classical assumption is to require that $\mathcal{X}$ is of type $p > 1$, *ie* there is $b > 0$ such that for every sequence $(Y_i)_i$ of centered random variables with $\mathbb{E}\left[\|Y_i\|^p\right] < +\infty$, one has

$$\mathbb{E}\left[\|Y_1 + \cdots + Y_n\|^p\right] \leq b\left[\mathbb{E}\left[\|Y_1\|^p\right] + \cdots + \mathbb{E}\left[\|Y_n\|^p\right]\right]. \tag{6.20}$$

If $\mathcal{X}$ is of type p and $\mathbb{E}\left[\|X_1\|^p\right] < +\infty$, then one can deduce immediately from (6.20) the following control:

$$\mathbb{E}\left[Z_n\right] \leq \frac{1}{n^{1-1/p}}\left(b\mathbb{E}\left[\|X_1 - \mathbb{E}[X_1]\|^p\right]\right)^{1/p}. \tag{6.21}$$

Controls like (6.21) can be used in Theorem 6.12 to derive precise deviations bounds for empirical means. Let us conclude this section with a concrete example.

Theorem 6.22. *Let μ be a probability measure on a Banach space $(\mathcal{X}, \|\cdot\|)$ such that $\int_{\mathcal{X}} e^{a\|x\|}\,\mu(dx) < +\infty$, for some $\delta > 0$. Then, for all sequence X_i of iid random variables with law μ, one has*

$$\mathbb{P}\left(Z_n \geq \mathbb{E}[Z_n] + t\right) \leq e^{-n\left(\sqrt{1+\frac{t}{M}}-1\right)^2}, \quad \forall t > 0, \tag{6.23}$$

where Z_n is defined by (6.19) and $M := \inf\left\{b > 0 : \iint_{\mathcal{X}^2} e^{\frac{\|x-y\|}{b}}\mu(dx)\mu(dy) \leq 2\right\}$.

Proof. According to Corollary 3.26, μ satisfy the T_1-inequality

$$\alpha\left(\mathcal{T}_{\|\cdot\|}(\mu,\nu)\right) \leq H(\nu \mid \mu), \quad \forall \nu \in \mathcal{P}(\mathcal{X}),$$

with $\alpha(t) = \left(\sqrt{1+\frac{t}{M}}-1\right)^2$. Thus, applying Theorem 6.12, the result follows immediately. □

Inequality (6.23) is very close to a well known inequality by Yurinskii ([18], Theorem 2.1). Under the same assumptions on μ, one can easily derive from Yurinskii's result the following bound :

$$\mathbb{P}\left(Z_n \geq \mathbb{E}\left[Z_n\right]+t\right) \leq \exp\left(-\frac{1}{8}\frac{nt^2}{2M_0^2+tM_0}\right), \quad \forall t>0, \tag{6.24}$$

where $M_0 = \inf\left\{b>0 : \int_{\mathcal{X}} e^{\frac{\|x\|}{b}}\,\mu(dx) \leq 2\right\}$. To compare (6.23) and (6.24) first note that

$$\left(\sqrt{1+u}-1\right)^2 \geq \frac{u^2}{2(2+u)}, \quad \forall u>0, \tag{6.25}$$

(this is left to the reader). Next, let us show that

$$M \leq 2M_0. \tag{6.26}$$

This follows from the following inequality :

$$\iint_{\mathcal{X}^2} e^{\frac{\|x-y\|}{2M_0}}\,\mu(dx)\mu(dy) \overset{(i)}{\leq} \left(\int_{\mathcal{X}} e^{\frac{\|x\|}{2M_0}}\,\mu(dx)\right)^2 \overset{(ii)}{\leq} \int_{\mathcal{X}} e^{\frac{\|x\|}{M_0}}\,\mu(dx) \overset{(iii)}{\leq} 2,$$

where (i) comes from the triangle inequality, (ii) from Jensen inequality and (iii) from the definition of M_0. Thanks to (6.25) and (6.26), one obtains

$$\left(\sqrt{1+\frac{t}{M}}-1\right)^2 \geq \frac{t^2}{2(2M^2+tM)} \geq \frac{t^2}{8(2M_0^2+tM_0/2)} \geq \frac{t^2}{8(2M_0^2+tM_0)}.$$

Thus, (6.23) is a little bit stronger than (6.24).
Yurinskii's proof relies on martingale arguments, while our proof is a direct consequence of the tensorization mechanism.

7. Large deviations and $\mathcal{T}$-inequalities. Abstract results

The framework is the same as in Section 5. See in particular Remark 5.4.

7.1. **A deviation function is a transportation function.** In this section, we give a rigorous proof at Theorem 7.1 of the Recipe 2.8 for an increasing deviation function which may possibly be not convex. This extends Proposition 5.5.

Theorem 7.1. *Let us assume (3.5) and (3.6).*

(a) *Any deviation function is a transportation function.*
(b) *If in addition $\mathcal{T}$ is continuous on $\mathcal{P}_{\mathcal{F}}$, then the converse also holds: any transportation function is a deviation function.*

Proof. (a) As $\mathcal{T}$ is lower semicontinuous, for all $t \geq 0$ the set $\{\nu \in \mathcal{P}_\mathcal{F}; \mathcal{T}(\nu) > t\}$ is open. It follows with the LD lower bound that

$$-\inf\{H(\nu \mid \mu); \nu \in \mathcal{P}_\mathcal{F}, \mathcal{T}(\nu) > t\} \leq \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\mathcal{T}(L_n) > t)$$

Let α be any deviation function: for all $t \geq 0$, $\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\mathcal{T}(L_n) \geq t) \leq -\alpha(t)$. Hence we obtain $\alpha(t) \leq \inf\{H(\nu \mid \mu); \nu \in \mathcal{P}_\mathcal{F}, \mathcal{T}(\nu) > t\}$ so that $\alpha(t-\delta) \leq H(\nu \mid \mu)$ for all $\nu \in \mathcal{P}_\mathcal{F}$ and $\delta > 0$ such that $\mathcal{T}(\nu) > t - \delta$. Taking $t = \mathcal{T}(\nu)$ leads us to $\alpha(\mathcal{T}(\nu) - \delta) \leq H(\nu \mid \mu)$ for all $\nu \in \mathcal{P}_\mathcal{F}$ and $\delta > 0$. As α is increasing and $\delta > 0$ is arbitrary, we have $\alpha(\mathcal{T}(\nu)^-) \leq H(\nu \mid \mu)$. The desired result follows from the assumed left continuity of α.

(b) As $\mathcal{T}$ is continuous, because of the contraction principle, $\{\mathcal{T}(L_n)\}$ obeys the LDP with rate function $i(t) = \inf\{H(\nu \mid \mu); \nu \in \mathcal{P}_\mathcal{F}, \mathcal{T}(\nu) = t\}, t \geq 0$. In particular, the LD upper bound: $\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\mathcal{T}(L_n) \geq t) \leq -\inf\{i(s); s \geq t\}$, is satisfied.
Let α be a transportation function. It clearly satisfies $\alpha(t) \leq \inf\{H(\nu \mid \mu); \nu \in \mathcal{P}_\mathcal{F}, \mathcal{T}(\nu) = t\}$ for all t. That is: $\alpha \leq i$. Finally, for all $t \geq 0$,

$$\begin{aligned}\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\mathcal{T}(L_n) \geq t) &\leq -\inf_{s \geq t} i(s)\\ &\leq -\inf_{s\geq t} \alpha(s)\\ &= -\alpha(t)\end{aligned}$$

where the last equality holds because α is increasing. This means that α is a deviation function. □

Remarks.

- Note that we didn't use the specific form (2.4) of $\mathcal{T}$, but only its lower semicontinuity.
- Similarly, we didn't use the specific properties of the relative entropy, but only that it is a LDP rate function for $\{L_n\}$.
- Statement (b) will not be used later, but it is satisfactory to know that a transportation function is not far from being a deviation function. A natural situation where $\mathcal{T}$ is continuous appears with $c = d^p$ since the Wasserstein's metric $\mathcal{T}_{d^p}^{1/p}$ metrizes $\sigma(\mathcal{P}_\mathcal{F}, \mathcal{F})$ with $\mathcal{F}$ the space of all continuous functions φ such that $|\varphi(x)| \leq c(1 + d(x_o, x)^p), \forall x$ for some constant c, see ([17], Chapter 7).

7.2. **The transportation function J_Φ.** With Theorem 7.1 in hand, it is enough to compute a deviation function α to obtain the TCI

$$\alpha(\mathcal{T}(\nu)) \leq H(\nu \mid \mu), \quad \forall \nu \in \mathcal{P}_\mathcal{F} \tag{7.2}$$

But these functions may be rather hard to compute because of the sup in the definition (2.4) of

$$T_n = \mathcal{T}(L_n) = \sup_{(\psi,\varphi)\in\Phi} \{\langle \varphi, L_n\rangle + \langle \psi, \mu\rangle\}.$$

However, it is shown at Theorem 7.7 below, that more can be said about transportation functions.

Assumptions (A). *The following requirements are assumed to hold.*

(i) *We assume (3.5):*

$$\int_{\mathcal{X}} e^{\varphi}\, d\mu < \infty, \forall \varphi \in \mathcal{F}.$$

(ii) *We assume (3.6):*

$$(0,0) \in \Phi \subset \mathcal{F} \times \mathcal{F},$$

(iii) *For all* $(\psi, \varphi) \in \Phi$, $\psi + \varphi \leq 0$.

Requirement (iii) always holds in the norm case: $\Phi = \Phi_U$, and it holds in the transportation case $\Phi = \Phi_c$ if $c(x, x) = 0, \forall x \in \mathcal{X}$.
Let us define

$$\Lambda(\varphi) := \log \int_{\mathcal{X}} e^{\varphi}\, d\mu.$$

Proposition 7.3. *Under the assumption (3.5)*

(a) $\{L_n\}$ *obeys the LDP in* $\mathcal{P}_{\mathcal{F}}$ *with the rate function*

$$H(\nu \mid \mu) = \Lambda^*(\nu) = \sup_{\varphi \in \mathcal{F}} \{\langle \varphi, \nu \rangle - \Lambda(\varphi)\}, \quad \nu \in \mathcal{P}_{\mathcal{F}}. \tag{7.4}$$

(b) *and for all* $(\psi, \varphi) \in \Phi$, $\{T_n^{\psi,\varphi}\}_{n \geq 1}$ *obeys the LDP in* $\mathbb{R}$ *with the rate function*

$$J_{\psi,\varphi}(t) = \sup_{s \in \mathbb{R}} \{st - \Lambda(s\varphi) - s\langle \psi, \mu \rangle\},\ t \in \mathbb{R}.$$

Proof. Statement (a) is Theorem 3.2.
The function $J_{\psi,\varphi}$ is the convex conjugate of

$$\Lambda_{\psi,\varphi}(s) := \Lambda(s\varphi) + s\langle \psi, \mu \rangle,\ s \in \mathbb{R}.$$

Since $\Lambda_{\psi,\varphi}$ is a steep function under assumptions (ii) and (iii), (b) is a direct consequence of Gärtner-Ellis theorem. □

We know that $J_{\psi,\varphi}$ is convex with a minimum value 0 attained at $\Lambda'_{\psi,\varphi}(0)$. Under assumption (iii), we have $\Lambda'_{\psi,\varphi}(0) = \langle \varphi + \psi, \mu \rangle \leq 0$. Therefore, $J_{\psi,\varphi}$ is an increasing nonnegative function on $[0, \infty)$ and so are J_Φ and $\widetilde{J}_\Phi$ given by

$$\begin{aligned} J_\Phi(t) &:= \widetilde{J}_\Phi(t^-), t > 0 \quad \text{where} \\ \widetilde{J}_\Phi(t) &:= \inf_{(\psi,\varphi) \in \Phi} J_{\psi,\varphi}(t) \in [0, \infty], t \geq 0 \end{aligned} \tag{7.5}$$

with $J_\Phi(0) = 0$. This last equality follows from assumption (ii). As $\Lambda'_{\psi,\varphi}(0) \leq 0$, it also holds that for all $t \geq 0$, $J_{\psi,\varphi}(t) = \Lambda^{\circledast}_{\psi,\varphi}(t) := \sup_{s \geq 0}\{st - \Lambda_{\psi,\varphi}(s)\}$ where the sup is taken over $s \geq 0$ rather than $s \in \mathbb{R}$. It follows that one can equivalently define J_Φ as follows.

Definition 7.6 (of the functions J_Φ and J)**.** .

- J_Φ is the left continuous version of the increasing function

$$t \in [0, \infty) \mapsto \inf_{(\psi,\varphi) \in \Phi} \sup_{s \geq 0} \{st - \Lambda(s\varphi) - s\langle \psi, \mu \rangle\} \in [0, \infty].$$

- J is the *best transportation function*. Clearly, it is the left continuous function of the increasing function

$$t \in [0, \infty) \mapsto \inf\{H(\nu \mid \mu); \nu \in \mathcal{P}_{\mathcal{F}} : \mathcal{T}(\nu) \geq t\} \in [0, \infty].$$

Although the best transportation function J might be out of reach in many situations, we have the following reassuring result.

Theorem 7.7. *Suppose that Assumptions (A) hold. Then, J_Φ is a transportation function and the* best *transportation function* in the class $\mathcal{C}$ *is the convex lower semicontinuous regularization of J_Φ.*

Proof. This statement is a collection of the statements of Theorem 7.8-a and Corollary 7.11-a,b which will be proved below. □

Theorem 7.8. *Suppose that Assumptions (A) hold.*

(a) *Then, J_Φ is a transportation function for $\mathcal{T}$ and $\{L_n\}$. This can be equivalently rewritten as the following TCI*

$$J_\Phi(\mathcal{T}(\nu)) \le H(\nu \mid \mu), \ \forall \nu \in \mathcal{P}_\mathcal{F}.$$

(b) *If in addition $\mathcal{T}$ is continuous on $\mathcal{P}_\mathcal{F}$, then J_Φ is the best transportation function. It is also the best deviation function: This means that $J_\Phi = J$.*

Proof. (a) As $\nu \mapsto \langle \varphi, \nu\rangle + \langle \psi, \mu\rangle$ is continuous, it follows from the contraction principle that $J_{\psi,\varphi}(t) = \inf\{H(\nu \mid \mu); \nu \in \mathcal{P}_\mathcal{F}, \langle \varphi, \nu\rangle + \langle \psi, \mu\rangle = t\}$ for all $t \ge 0$. Hence, $J_{\psi,\varphi}(\langle \varphi, \nu\rangle + \langle \psi, \mu\rangle) \le H(\nu \mid \mu)$ for all $\nu \in \mathcal{P}_\mathcal{F}$ and a fortiori

$$\widetilde{J}_\Phi(\langle \varphi, \nu\rangle + \langle \psi, \mu\rangle) \le H(\nu \mid \mu),$$

as soon as $\langle \varphi, \nu\rangle + \langle \psi, \mu\rangle \ge 0$. As $\widetilde{J}_\Phi$ is increasing, by the definition (2.4) of $\mathcal{T}(\nu)$, one obtains: $\widetilde{J}_\Phi(\mathcal{T}(\nu)^-) \le H(\nu \mid \mu)$ which is the desired result. Note that $\mathcal{T}(\nu) \ge 0$ since $(0,0) \in \Phi$ (assumption (A.ii)).

(b) Because of part (b) of Theorem 7.1, it is enough to prove that $J_\Phi = J$. Because of part (a) of the present theorem, J_Φ is a transportation function, and by part (b) of Theorem 7.1, it is also a deviation function. Therefore, $J_\Phi \le J$ and it remains to prove that $J \le J_\Phi$.
By the LD lower bound for $\{T_n^{\psi,\varphi}\}$, for all $t \ge 0$,

$$\begin{aligned}
-\inf_{r>t} J_{\psi,\varphi}(r) &\le \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T_n^{\psi,\varphi} > t) \\
&\le \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\sup_{(\psi,\varphi)\in\Phi} T_n^{\psi,\varphi} \ge t \right) \\
&\le -J(t).
\end{aligned}$$

Since $J_{\psi,\varphi}$ is increasing, we have: $J(t) \le \inf_{r>t} J_{\psi,\varphi}(r) = J_{\psi,\varphi}(t^+)$, so that for all $t \ge 0$

$$\begin{aligned}
J(t) &\le \inf\{J_{\psi,\varphi}(t^+), (\psi,\varphi) \in \Phi\} \\
&= \inf_\phi \inf_{u>t} J_\phi(u) \\
&= \inf_{u>t} \inf_\phi J_\phi(u) \\
&= \widetilde{J}_\Phi(t^+).
\end{aligned}$$

As J and $\widetilde{J}_\Phi$ are increasing and J is left continuous, this gives $J(t) \le \widetilde{J}_\Phi(t^-)$ for all $t > 0$ which is the desired result. □

7.3. **Connections with Theorem 3.7.** Let us first give an alternative proof of criterion $(b) \Rightarrow (a)$ of Theorem 3.7.

We keep the Assumptions (A) of Section 7.2. Note that because of Assumptions (A.ii) and (A.iii), the function

$$\Lambda_\Phi(s) := \sup_{(\psi,\varphi)\in\Phi} \Lambda_{\psi,\varphi}(s) = \sup_{(\psi,\varphi)\in\Phi} \{\Lambda(s\varphi) + s\langle\psi,\mu\rangle\}, \ s \geq 0 \tag{7.9}$$

is in the class $\mathcal{C}$. It follows that its monotone conjugate

$$\Lambda_\Phi^\circledast(t) = \sup_{s\geq 0}\{st - \Lambda_\Phi(s)\}, t \geq 0$$

is also in $\mathcal{C}$. Thanks to formula (7.5), for all $t \geq 0$, we have

$$\begin{aligned}
\Lambda_\Phi^\circledast(t) &\leq \sup_{s\geq 0}\left\{ st - \sup_{(\psi,\varphi)\in\Phi} \Lambda_{\psi,\varphi}(s)\right\} \\
&= \sup_{s\geq 0} \inf_{(\psi,\varphi)\in\Phi} \{st - \Lambda_{\psi,\varphi}(s)\} \\
&\leq \inf_{(\psi,\varphi)\in\Phi} \sup_{s\geq 0}\{st - \Lambda_{\psi,\varphi}(s)\} \\
&= \widetilde{J}_\Phi(t)
\end{aligned}$$

But $\Lambda_\Phi^\circledast(t)$ is left continuous, hence

$$\Lambda_\Phi^\circledast \leq J_\Phi. \tag{7.10}$$

As J_Φ is a transportation function (Theorem 7.8), so is $\Lambda_\Phi^\circledast$.

The criterion $(b) \Rightarrow (a)$ of Theorem 3.7 follows from the above considerations. Indeed, (b) states that $\Lambda_\Phi \leq \alpha^\circledast$. Therefore, with (7.10): $\alpha \leq \Lambda_\Phi^\circledast \leq J_\Phi$. Hence, α is a transportation function.

An easy consequence of Theorem 3.7 is the following

Corollary 7.11. *Suppose that Assumptions (A) hold.*

- (a) *The best transportation function in the class $\mathcal{C}$ is $\Lambda_\Phi^\circledast$. This means that $\alpha \in \mathcal{C}$ is a transportation function if and only if $\alpha \leq \Lambda_\Phi^\circledast$.*
- (b) *Moreover, $\Lambda_\Phi^\circledast$ is the convex lower semicontinuous regularization of J_Φ (in restriction to $t \in [0,\infty)$).*
- (c) *If $\mathcal{T}$ is continuous, then $\Lambda_\Phi^\circledast$ is also the best deviation function in the class $\mathcal{C}$.*

Proof. The best function $\alpha^\circledast \in \mathcal{C}$ satisfying (b) of Theorem 3.7 is $\alpha^\circledast = \Lambda_\Phi$, see (7.9). Because of the equivalence $(a) \Leftrightarrow (b)$ of Theorem 3.7, its monotone conjugate $\Lambda_\Phi^\circledast$ is the best transportation function in $\mathcal{C}$. This is (a).

Let us prove (b). In order to work with usual convex conjugates, let us state $J_\phi(t) = +\infty$ for all $t < 0$ and $\phi \in \Phi$. We have

$$\begin{aligned}
(\inf_\phi J_\phi)^*(s) &= \sup_t\{st - \inf_\phi J_\phi(t)\} \\
&= \sup_{t,\phi}\{st - J_\phi(t)\} \\
&= \sup_\phi \sup_t\{st - J_\phi(t)\} \\
&= \sup_\phi J_\phi^*(s).
\end{aligned}$$

Hence, the convex lower semicontinuous regularization of $J_\Phi := \inf_\phi J_\phi$ is $(\inf_\phi J_\phi)^{**} = (\sup_\phi J_\phi^*)^* = (\sup_\phi \Lambda_\phi^{**})^*$ But, the convex lower semicontinuous regularization of $\sup_\phi \Lambda_\phi$ is $\sup_\phi \Lambda_\phi^{**}$. Therefore, $J_\Phi^{**} = (\sup_\phi \Lambda_\phi^{**})^* = (\sup_\phi \Lambda_\phi)^* = \Lambda_\Phi^*$. But it is already seen that in restriction to $t \in [0, \infty)$, Λ_Φ is in $\mathcal{C}$, so that $\Lambda_\Phi^*(t) = \Lambda_\Phi^{\circledast}(t)$ for all $t \geq 0$.
Finally, (c) is a direct consequence of (b) and Theorem 7.8-(b). □

References

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

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

[3] F. Bolley, A. Guillin, and C. Villani. Quantitative concentration inequalities for empirical measures on non-compact spaces. Preprint.
Available online via `http://www.ceremade.dauphine.fr/~guillin/index3.html`, 2005.

[4] I. Csiszár. *I*-divergence geometry of probability distributions and minimization problems. *Annals of Probability*, 3:146–158, 1975.

[5] I. Csiszár. Sanov property, generalized *I*-projection and a conditional limit theorem. *Annals of Probability*, 12:768–793, 1984.

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

[7] 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.

[8] P. Eichelsbacher and U. Schmock. Large deviations of U-empirical measures in strong topologies and applications. *Annales de l'Institut Henri Poincaré*, 38(5):779–797, 2002.

[9] N. Gozlan. *Principe conditionnel de Gibbs pour des contraintes fines approchées et Inégalités de Transport.* PhD thesis, Université de Paris 10, 2005.

[10] M. Ledoux. *The Concentration of Measure Phenomenon.* Mathematical Surveys and Monographs 89. American Mathematical Society, Providence RI, 2001.

[11] C. Léonard and J. Najim. An extension of Sanov's theorem : application to the Gibbs conditioning principle. *Bernoulli*, 8(6):721–743, 2002.

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

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

[14] P. Massart. Saint-Flour Lecture Notes, 2003. concentration inequalities and model selection.
Available online via `http://www.math.u-psud.fr/~massart/`.

[15] S. Rachev and L. Rüschendorf. *Mass Transportation Problems. Vol I : Theory, Vol. II : Applications.* Probability and its applications. Springer Verlag, New York, 1998.

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

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

[18] V.V. Yurinskii. Exponential inequalities for sums of random vectors. *Journal of multivariate analysis*, 6:473–499, 1976.

(Nathael Gozlan) Modal-X, Université Paris 10. Bât. G, 200 av. de la République. 92001 Nanterre Cedex, France
E-mail address: `nathael.gozlan@u-paris10.fr`

(Christian Léonard) Modal-X, Université Paris 10. Bât. G, 200 av. de la République. 92001 Nanterre Cedex, France

(Christian Léonard) CMAP, École Polytechnique. 91128 Palaiseau Cedex, France
E-mail address: christian.leonard@polytechnique.fr