Probab. Theory Relat. Fields (2007) 139:235–283
DOI 10.1007/s00440-006-0045-y

A large deviation approach to some transportation cost inequalities

Nathael Gozlan · Christian Léonard

Received: 18 July 2005 / Revised: 6 November 2006 / Published online: 12 December 2006
© Springer-Verlag 2006

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.

Keywords Transportation cost inequalities · Large deviations · Concentration of measure

Mathematics Subject Classification (2000) 60E15 · 60F10

Contents

N. Gozlan · C. Léonard
Modal-X, Université Paris 10, Bât. G, 200 av. de la République, 92001 Nanterre Cedex, France
e-mail: nathael.gozlan@univ-mlv.fr

C. Léonard (✉)
CMAP, École Polytechnique, 91128 Palaiseau Cedex, France
e-mail: christian.leonard@polytechnique.fr

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

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

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

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*

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

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

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

if ν is absolutely continuous with respect to μ and $H(\nu \mid \mu) = \infty$ otherwise. In presence of the family of inequalities (1), one says that μ satisfies $T_p(C)$.

For instance, Csiszár–Kullback–Pinsker's inequality, see (18), is $T_1(1)$ with 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 Otto and Villani [19] and by Bobkov et al. [2], 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 [22], 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 $T_1(C)$, Marton [16,17] has obtained the following *concentration inequality* for μ:

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

for all measurable subset A such that $\mu(A) \geq 1/2$ and all $r \geq \sqrt{2C\log 2}$. *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 Marton [17] (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 the inequality $T_1(C_1 + \cdots + C_n)$ (for the cost function $\sum_{i=1}^n d(x_i, y_i)$), while when $p = 2$, $\mu_1 \otimes \cdots \otimes \mu_n$ satisfies $T_2(\max(C_1, \dots, C_n))$ (for the cost function $\sum_{i=1}^n d^2(x_i, y_i)$). 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 so does the infinite product $\mu^{\otimes\infty}$. One says that T_2-inequalities have the *dimension-free tensorization property*. This property was first established by Talagrand in [22].

By Jensen's inequality, we have $(\mathcal{T}_d)^2 \leq \mathcal{T}_{d^2}$ so that $T_2(C)$ implies $T_1(C)$. As the standard normal law γ satisfies $T_2(1)$, it follows easily from the dimension-free tensorization property that the standard normal law on $\mathbb{R}^n$, denoted by γ^n, satisfies $T_2(1)$ and therefore $T_1(1)$ and the concentration inequality

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

for all measurable subset A such that $\mu(A) \geq 1/2$ and all $r \geq \sqrt{2\log 2}$ 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 which is: $\gamma^n(\{x; d(x,A) > r\}) \leq \frac{1}{\sqrt{2\pi}} \int_r^{+\infty} e^{-u^2/2}\, \mathrm{d}u$, for all $r \geq 0$, see Borell's paper [6] or Ledoux's monograph [13], p. 28. In view of (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 (2) holds for some C.

Criteria for T_1 It has recently been proved by Djellout et al. in [10] that μ satisfies $T_1(C)$ for some C if and only if

$$\int_{\mathcal{X}} e^{a_o d(x_o,x)^2}\, \mu(\mathrm{d}x) < \infty \tag{3}$$

for some $a_o > 0$ and some (and therefore all) x_o in $\mathcal{X}$. It follows that (3) *is a characterization of the Gaussian concentration*. The proof of this result in [10] relies on a dual characterization of T_1 which has been obtained by Bobkov and

Götze in [1]. This characterization is the following: $T_1(C)$ holds if and only if

$$\log \int_{\mathcal{X}} e^{s(\varphi - \langle \varphi, \mu \rangle)} \, \mathrm{d}\mu \leq Cs^2/2, \tag{4}$$

for all $s \geq 0$ and all bounded Lipschitz function φ with $\|\varphi\|_{\mathrm{Lip}} \leq 1$, where $\langle \varphi, \mu \rangle$ will denote in the sequel the usual duality bracket between measures and functions, that is $\langle \varphi, \mu \rangle := \int_{\mathcal{X}} \varphi \, \mathrm{d}\mu$.

The criterion (3) has been recovered very recently by Bolley and Villani in [5] 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 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) enter the following larger class of inequalities, which will also be called transportation cost inequalities (TCIs):

$$\alpha(\mathcal{T}_c(\mu, \nu)) \leq H(\nu \mid \mu), \quad \forall \nu \in \mathcal{P}(\mathcal{X}) \tag{5}$$

where $\alpha : [0, \infty) \to [0, \infty)$ is an increasing[1] function which vanishes at 0. The inequality (1) corresponds $c = d^p$ with $\alpha(t) = t^{2/p}/(2C), t \geq 0$. Of course, one should rigorously restrict (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

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

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

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

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 Sect. 2.2. We obtain the

Recipe 1 *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 (5).

Rigorously, one will have to require that α is a left continuous function. This result will be proved at Theorem 15 and a weak version of it (with α convex) is proved at Proposition 4. Under certain hypotheses, one can show that the left continuous version of the function

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

is optimal in (5), see Corollary 9 at Sect. 7.

Not only TCIs can be derived with this recipe but also another class of functional inequalities which we call Norm-Entropy Inequalities (NEIs), see (11) 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 Sect. 6.

As regards item (iii), the main technical (easy) result is Theorem 2 which is an extension of Bobkov and Götze's characterization of $T_1(C)$ stated at (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 [5] about weighted CKP inequalities at Corollary 3.

Tensorization of convex TCIs is also handled. The main result on this topic is Theorem 5. 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 Sect. 5. It emerges from our analysis via large deviations, that integral criteria only control the behavior of $\alpha(t)$ in (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 7. It is the following:

Let d be a lower semicontinuous metric. Suppose that $a \geq 0$ satisfies $\int_{\mathcal{X}} e^{ad(x_o,x)}\, \mu(\mathrm{d}x) \leq 2$ for some $x_o \in \mathcal{X}$ and that γ is an increasing convex function which satisfies $\gamma(0) = 0$ and $\int_{\mathcal{X}} e^{\gamma(d(x_1,x))}\, \mu(\mathrm{d}x) \leq B < \infty$ for some $x_1 \in \mathcal{X}$, then

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

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 (3) of [10] and [5].

The last Sect. 7 is devoted to abstract results. In particular, the extended version Recipe 2 of Recipe 1 is proved at Theorem 15. 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.

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

$$\mathcal{T}_c(\mu,\nu) = \sup_{(\psi,\varphi)\in\Phi_c} \left\{ \int_{\mathcal{X}} \psi\, \mathrm{d}\mu + \int_{\mathcal{X}} \varphi\, \mathrm{d}\nu \right\} \tag{6}$$

where Φ_c is the set of all couples (ψ,φ) of Borel measurable bounded functions on $\mathcal{X}$ such that $\psi(x) + \varphi(y) \leq c(x,y)$ 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 (a proof of this well known result can be found in Chap. 1 of [23]). 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,

$$\mathcal{T}_d(\mu,\nu) = \sup\left\{ \int_{\mathcal{X}} \varphi\, \mathrm{d}(\nu-\mu); \varphi \in B(\mathcal{X}), \|\varphi\|_{\mathrm{Lip}} \leq 1 \right\} := \|\nu-\mu\|^*_{\mathrm{Lip}} \tag{7}$$

where the space of all Borel measurable bounded functions on $\mathcal{X}$ is denoted $B(\mathcal{X})$ and $\|\varphi\|_{\text{Lip}} = \sup_{x\neq y} \frac{|\varphi(x)-\varphi(y)|}{d(x,y)}$ 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 $\|\nu-\mu\|^*_{\text{Lip}}$ (for a proof, see Chap. 1 of [23]).

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

After these considerations, it appears that the transportation cost inequality (1) enters the following larger class of inequalities, which we call $\mathcal{T}$-inequalities:

$$\alpha(\mathcal{T}(\nu)) \le H(\nu \mid \mu), \quad \forall \nu \in \mathcal{N} \tag{8}$$

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

$$\mathcal{T}(\nu) = \sup_{(\psi,\varphi)\in\Phi} \left\{ \int_{\mathcal{X}} \psi \, \mathrm{d}\mu + \int_{\mathcal{X}} \varphi \, \mathrm{d}\nu \right\} \tag{9}$$

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 (8) 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 correspond 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 continuous bounded functions on $\mathcal{X}$ is denoted $C_b(\mathcal{X})$. In the situation where Φ is equal to

$$C_b \cap \Phi_c := \{(\psi,\varphi) \in C_b(\mathcal{X}) \times C_b(\mathcal{X}); \psi(x) + \varphi(y) \le c(x,y), \forall x, y \in \mathcal{X}\}$$

the family of inequalities (8) is called a transportation cost inequality (TCI). Indeed, the Kantorovich dual equality (6) 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 (8) is

$$\alpha(\mathcal{T}_c(\mu,\nu)) \le H(\nu \mid \mu), \quad \forall \nu \in \mathcal{N} \tag{10}$$

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 $\int_{\mathcal{X}} \chi \, \mathrm{d}\nu < \infty$.

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

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

In this case, inequality (8) is

$$\alpha(\|\nu - \mu\|_U^*) \leq H(\nu \mid \mu), \quad \forall \nu \in \mathcal{P}_U \tag{11}$$

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 (11) 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 $U := \{\varphi \in F, \|\varphi\| \leq 1\}$ 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 (7)) 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 of 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 Sects. 3 and 7, $\mathcal{T}$-inequalities will be proved by means of a large deviation approach. A good reference for large deviations theory is the book of Dembo and Zeitouni [9]. 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

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

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 $t \geq 0$

$$\limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(T_n \geq t) \leq -\alpha(t). \tag{12}$$

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 (8). We have just obtained the correct reformulation of Recipe 1:

Recipe 2 *Any deviation function α of $\{T_n\}$ satisfies the $\mathcal{T}$-inequality* (8).

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 2, Proposition 4 and Theorem 15.

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

In the rest of the paper (except Sect. 7) our attention is restricted to those $\mathcal{T}$-inequalities (8) where the function α is increasing and convex. In this case, (8) 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 2.

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

$$\mathcal{F}_{\exp}(\mu) = \left\{ \varphi : \mathcal{X} \to \mathbb{R}; \varphi \text{ measurable}, \int_{\mathcal{X}} \exp(a|\varphi|)\, \mathrm{d}\mu < \infty \text{ for all } a > 0 \right\} \tag{13}$$

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

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

the set of all probability measures which integrate every function of $\mathcal{F}_{\exp}(\mu)$.

The set $\mathcal{N}_{\exp}(\mu)$ is furnished with the cylinder σ-field generated by the functions $\nu \mapsto \langle \varphi, \nu \rangle, \varphi \in \mathcal{F}_{\exp}(\mu)$ and is endowed with the topology $\sigma(\mathcal{N}_{\exp}(\mu), \mathcal{F}_{\exp}(\mu))$, that is, the coarsest topology which makes the maps $\nu \mapsto \langle \varphi, \nu \rangle$ continuous for all $\varphi \in \mathcal{F}_{\exp}(\mu)$.

Theorem 1 (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

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

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 [9], Theorem 6.2.10). For two distinct detailed proofs of the present theorem, see ([11], Theorem 1.7) or ([15], Corollary 3.3). For the original result by Sanov, see [21].

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 1 (of $\mathcal{C}$) *The class $\mathcal{C}$ consists of all the $[0,\infty]$-valued 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 only function $\alpha \in \mathcal{C}$ which is not right continuous at 0 satisfies $\alpha(t) = \infty$ for all $t > 0$.

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 1 *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 2 below is a criterion for a convex $\mathcal{T}$-inequality to hold. It extends two well-known results of Bobkov and Götze ([1], Theorem 1.3 and statement (1.7)).

Let $\mathcal{F}$ be a vector space of measurable functions φ on $\mathcal{X}$ such that

$$\int_{\mathcal{X}} e^{\varphi}\, \mathrm{d}\mu < \infty, \quad \forall \varphi \in \mathcal{F}. \tag{14}$$

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

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

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

$$(0,0) \in \Phi \subset \mathcal{F} \times \mathcal{F}, \tag{15}$$

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,

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

Theorem 2 *We assume* (14) *and* (15). *Let us consider the following statements where* α *is any function in* $\mathcal{C}$:

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

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

$$\int_{\mathcal{X}} (\varphi(x) + \psi(x)) \, \mu(\mathrm{d}x) \le 0 \tag{16}$$

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 (16) holds for all NEIs. It also holds for TCIs under the assumption that c satisfies

$$c(x,x) = 0, \quad \forall x \in \mathcal{X}. \tag{17}$$

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

Example 1 (CKP inequality and Hoeffding's inequality) Let $\Phi := \{(\varphi, -\varphi) : \varphi$ measurable such that $|\varphi(x)| \le 1, \forall x \in \mathcal{X}\}$. The associated $\mathcal{T}(\nu)$ is the total variation distance between ν and μ, that is

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

According to Hoeffding's Lemma (see e.g. [18]),

$$\Lambda_\phi(s) = \log \int_{\mathcal{X}} e^{s(\varphi(x) - \langle \varphi, \mu \rangle)}\, \mu(\mathrm{d}x) \leq \frac{s^2}{2},$$

for all $\phi \in \Phi$ and $s \geq 0$. Letting $\alpha(t) = \frac{t^2}{2}$, one has $\alpha^{\circledast}(s) = \frac{s^2}{2}$. Thus, applying Theorem 2, one gets

$$\frac{1}{2}\|\nu - \mu\|_{\mathrm{TV}}^2 \leq H(\nu \mid \mu), \quad \forall \nu \in \mathcal{P}(\mathcal{X}). \tag{18}$$

This inequality is the celebrated CKP inequality.

Proof (of Theorem 2) 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 (14) and (15) 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 (14) is equivalent to $\mathcal{F} \subset \mathcal{F}_{\exp}(\mu)$. It follows that under this assumption, Sanov's Theorem 1 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 $n \geq 1$,

$$T_n^\phi = \langle \varphi, L_n \rangle + \langle \psi, \mu \rangle = \frac{1}{n} \sum_{i=1}^n (\varphi(X_i) + \mathbb{E}\psi(X_i)) \tag{19}$$

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

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

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

$$\begin{aligned} -\inf_{u > t} \Lambda_\phi^*(u) &\leq \liminf_{n \to \infty} \frac{1}{n} \log \mathbb{P}(T_n^\phi > t) \\ &\leq \limsup_{n \to \infty} \frac{1}{n} \log \mathbb{P}(T_n^\phi \geq t) \leq -\inf_{u \geq t} \Lambda_\phi^*(u) \end{aligned} \tag{20}$$

Because of assumption (15), 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

$$\Lambda_\phi^*(t) = \inf\{H(\nu \mid \mu); \nu \in \mathcal{P}_\mathcal{F} : \langle \varphi, \nu\rangle + \langle \psi, \mu\rangle = t\}. \tag{21}$$

It is convenient to extend α on $\mathbb{R}$ by taking $\alpha(t) = 0$, for all $t \le 0$. Doing so, one has : $\alpha^{\circledast}(s) = \alpha^*(s)$, for all $s \ge 0$ and $\alpha^*(s) = +\infty$, for all $s \le 0$, where α^* is the usual convex conjugate of α.

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

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

The equivalence (i) follows from the definition (9) of $\mathcal{T}$, (ii) holds true because α is increasing and left continuous while (iii) follows from (21).

$[(b) \Leftrightarrow (c)]$.

Let us prove $(c) \Rightarrow (b)$. As $\alpha(t) = 0$, for all $t \le 0$, statement (c) is equivalent to

$$\alpha(t) \le \Lambda_\phi^*(t), \quad \forall t \in \mathbb{R}, \ \forall \phi \in \Phi. \tag{22}$$

As, Λ_ϕ is convex and lower semicontinuous, we have: $\Lambda_\phi^{**} = \Lambda_\phi$. Hence, taking the convex conjugates on both sides of (22) one obtains that $\Lambda_\phi \le \alpha^*$ 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 $\alpha \le \Lambda_\phi^*$ 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 $t \le m$, we have $\inf_{u>t} \Lambda_\phi^*(u) = \inf_{u\ge t} \Lambda_\phi^*(u) = 0$. As Λ_ϕ^* is convex, it is continuous on (t_-, t_+) the interior of its effective domain. Therefore, we have for all $t \neq t_+, \inf_{u>t} \Lambda_\phi^*(u) = \inf_{u\ge t} \Lambda_\phi^*(u)$. Together with (20), this gives for all $t \neq t_+$,

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

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

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

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 (16). To obtain the stated series of equivalences, it remains to prove $(c)\Rightarrow(e)$.

By (19), $T_n^\phi=\frac{1}{n}\sum_{i=1}^n Y_i$ 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 $s\geq 0$,

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

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

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

But, assumption (16) 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 [23]) 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

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

without any change in the value of $\mathcal{T}_c$. One easily proves that if (17) 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, see [14].) 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 (14) and (15) hold with $\mathcal{F} = B(\mathcal{X})$.

Now, as a corollary of Theorem 2, we have the following result.

Corollary 1 *Whenever $\alpha \in \mathcal{C}$, the transportation cost inequality (10) holds in $\mathcal{N} = \mathcal{P}(\mathcal{X})$ if and only if*

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

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 2 specializes as follows.

Theorem 3 *Suppose that U satisfies*

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

Let α be in $\mathcal{C}$. Then, the norm-entropy inequality (11)

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

holds if and only if

$$\Lambda_\varphi(s) := \log \int_{\mathcal{X}} e^{s[\varphi(x) - \langle\varphi,\mu\rangle]} \, \mu(\mathrm{d}x) \leq \alpha^{\circledast}(s) \tag{23}$$

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

Specializing Theorem 3 by taking U to be the set of all 1-Lipschitz measurable bounded functions with respect to some measurable metric d, one obtains the following characterization of convex T_1-inequalities.

Theorem 4 (T_1-inequality) *Let d be a lower semicontinuous metric on $\mathcal{X}$ and α be in $\mathcal{C}$. Then,*

$$\alpha(\mathcal{T}_d(\mu, \nu)) \leq H(\nu \mid \mu),$$

for all $\nu \in \mathcal{P}(\mathcal{X})$ such that $\int_{\mathcal{X}} d(x_o, x) \, \nu(\mathrm{d}x) < \infty$ if and only if

$$\Lambda_\varphi(s) := \log \int_{\mathcal{X}} e^{s[\varphi(x) - \langle\varphi,\mu\rangle]} \, \mu(\mathrm{d}x) \leq \alpha^{\circledast}(s) \tag{24}$$

for all $s \geq 0$ *and all measurable bounded Lipschitz function* φ *such that* $\|\varphi\|_{\mathrm{Lip}} \leq 1$.

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

Proposition 2 *Assuming that* $\mathcal{F}$ *contains functions which are not* μ*-a.e. constant, the function* α *of a convex norm-entropy inequality* (11) *satisfies*

$$0 \leq \alpha(t) \leq at^2, \quad \forall 0 \leq t \leq t_1 \tag{25}$$

for some $a > 0$ *and* $t_1 > 0$.

Proof Let φ_o be a non constant function in U. Then, $\sigma_o^2 := \int_{\mathcal{X}} (\varphi(x) - \langle \varphi, \mu \rangle)^2 \, \mathrm{d}\mu > 0$ and for any $0 < \sigma_1^2 < \sigma_o^2$ there exists $s_1 > 0$ such that $\Lambda_{\varphi_o}(s) = \sigma_o^2 s^2/2 + o(s^2) \geq \sigma_1^2 s^2/2$, for all $0 \leq s \leq s_1$. Let $\theta_1(s)$ match with $\sigma_1^2 s^2/2$ on $[0, s_1]$ and be extended on $[s_1, \infty)$ by the tangent affine function of $s \mapsto \sigma_1^2 s^2/2$ at $s = s_1$. As Λ_{φ_o} is convex, we have $\theta_1(s) \leq \Lambda_{\varphi_o}(s)$ for all $s \geq 0$.

Together with (23), we obtain $\theta_1 \leq \alpha^{\circledast}$. Taking the monotone conjugates on both sides of this inequality provides us with

$$\alpha(t) \leq \theta_1^{\circledast}(t) = \begin{cases} t^2/(2\sigma_1^2), & \text{if } 0 \leq t \leq s_1 \sigma_1^2 \\ +\infty, & \text{if } t > s_1 \sigma_1^2 \end{cases}$$

from which the desired result follows.

To explore some consequences of Theorem 3 (see Corollaries 2 and 3 below) one needs the notion of Orlicz space associated with the exponential function. It appears that the space $\mathcal{F}_{\exp}(\mu)$ introduced at (13) is the Orlicz space

$$\left\{ \varphi : \mathcal{X} \to \mathbb{R};\ \text{measurable}, \int_{\mathcal{X}} \rho(a\varphi) \, \mathrm{d}\mu < \infty \text{ for all } a > 0 \right\}$$

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

$$\begin{aligned} \|\varphi\|_\rho &:= \inf \left\{ b > 0; \int_{\mathcal{X}} \rho\left(\frac{\varphi}{b}\right) \mathrm{d}\mu \leq 1 \right\} \\ &= \inf \left\{ b > 0; \int_{\mathcal{X}} e^{|\varphi|/b} \, \mathrm{d}\mu \leq 2 \right\} \end{aligned} \tag{26}$$

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

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

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 $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L_{\rho^*}(\mu)$.

Let us state a useful technical lemma.

Lemma 1 (A Bernstein type inequality) *For any measurable function φ such that $\int_{\mathcal{X}} e^{a_o|\varphi|} \, \mathrm{d}\mu < \infty$ 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 uniformly $\|\cdot\|_\rho$-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 $1 \geq \int_{\mathcal{X}} \rho(\varphi/\beta) \, \mathrm{d}\mu = \sum_{k \geq 1} \langle |\varphi|^k, \mu \rangle / (k! \beta^k)$. 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 $s \geq 0$. This completes the proof of the lemma.

We are now ready to prove some corollaries of Theorem 2. For any measurable function f in $L_{\rho^*}(\mu)$, let

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

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

Corollary 2 *For any probability measure ν which is absolutely continuous with respect to μ and such that $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L_{\rho^*}(\mu)$, we have*

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

Note that this is the NEI: $\alpha_1(\|\frac{\mathrm{d}\nu}{\mathrm{d}\mu} - 1\|_\rho^*) \le H(\nu \mid \mu)$, *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 1 applied with $M = 1$, (23) 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 Bolley and Villani in [5] with other constants.

Corollary 3 (Weighted CKP inequalities) *Let χ be a nonnegative function such that $\int_{\mathcal{X}} e^{a_o\chi}\,\mathrm{d}\mu < \infty$ 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 $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L_{\rho^*}(\mu)$, $\|\chi \cdot (\nu - \mu)\|_{\mathrm{TV}}$ 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 $\int_{\mathcal{X}} e^{a_1\chi}\,\mathrm{d}\mu = \infty$), one must be careful. It happens that

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

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 1 applied with $M = \|\chi\|_\rho$, (23) 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 1 It follows from Corollaries 2 and 3, 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 (18) 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

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

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

Corollary 4 *Suppose that there exist $a_o > 0$ and $x_o \in \mathcal{X}$ such that $\int_{\mathcal{X}} e^{a_o d(x_o,x)}\mu(\mathrm{d}x) < \infty$. Then, $\|d\|_{\rho,\mu^{\otimes 2}} = \inf\{b > 0; \int_{\mathcal{X}\times\mathcal{X}} e^{d(x,y)/b}\mu(\mathrm{d}x)\mu(\mathrm{d}y) \le 2\} < \infty$ 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 4. Here $U = \{\varphi : \|\varphi\|_{\mathrm{Lip}} \le 1, \sup|\varphi| < \infty\} \subset \mathcal{F}_{\exp}(\mu)$. Let us show that

$$\sup_{\varphi\in U} \|\varphi - \langle\varphi,\mu\rangle\|_\rho \le \|d\|_{\rho,\mu^{\otimes 2}}. \tag{27}$$

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

$$\exp\left[s\left(\varphi(x) - \int_{\mathcal{X}} \varphi(y)\,\mu(\mathrm{d}y)\right)\right] \leq \int_{\mathcal{X}} \exp[s(\varphi(x) - \varphi(y))]\,\mu(\mathrm{d}y)$$
$$\leq \int_{\mathcal{X}} \exp[sd(x,y)]\,\mu(\mathrm{d}y).$$

Hence, integrating with respect to $\mu(\mathrm{d}x)$, one obtains (27).

Thanks to Lemma 1 applied with $M = \|d\|_{\rho,\mu^{\otimes 2}}$, (24) holds as follows: $\Lambda_\varphi(s) \leq \alpha_M^{\circledast}(s) := (Ms)^2/(1-Ms)$. Taking the monotone conjugate, we obtain $\alpha_M(t) = (\sqrt{t/M+1}-1)^2$, which is the desired result.

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 (17).

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 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

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

Lemma 2 *Let α_1 and α_2 belong to the class $\mathcal{C}$. Then,*

(a) $\alpha_1 \square \alpha_2 \in \mathcal{C}$ *and*
(b) $(\alpha_1 \square \alpha_2)^{\circledast} = \alpha_1^{\circledast} + \alpha_2^{\circledast}$

Proof This simple exercise is left to the reader.

The main result of this section is the following theorem.

Theorem 5 (Tensorization) *Let c_1 and c_2 be two continuous nonnegative cost functions which satisfy* (17). *Suppose that the convex TCIs*

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

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

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

This statement is known in some cases (see for instance [16,22,2].) In [10], Djellout et al. have obtained a tensorization property for the classical T_1 inequality, which holds for non product measures.

There are two ways to prove tensorization properties for TCIs: a direct one, due to Marton, which is based on a coupling argument, and an indirect one, due to Ledoux, which makes use of the dual characterization of TCIs. The coupling method is, to our opinion, much more intuitively appealing, but it has the disadvantage of raising difficult measurability questions. The interested reader can consult the Chapter VI of [12], where this is discussed in details. The proof below is based upon the indirect dual approach, making use of the characterization of Corollary 1 and follows the line of proof of ([13], Proposition 1.19).

Proof (of Theorem 5) Recall that, provided that c is continuous nonnegative and satisfy (17), $Q^c\varphi(x) = \inf_{y \in \mathcal{X}}\{\varphi(y) + c(y,x)\}$ 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 1, the convex TCIs "$\alpha_1(\mathcal{T}_1) \leq H_1$" and "$\alpha_2(\mathcal{T}_2) \leq H_2$" which are supposed to hold are equivalent to

$$\int_{\mathcal{X}_1} e^{sQ_1\theta_1}\, \mathrm{d}\mu_1 = \exp(\alpha_1^{\circledast}(s) + s\langle \theta_1, \mu_1 \rangle), \quad \forall s \geq 0,\ \forall \theta_1 \in B(\mathcal{X}_1) \tag{28}$$

$$\int_{\mathcal{X}_2} e^{sQ_2\theta_2}\, \mathrm{d}\mu_2 = \exp(\alpha_2^{\circledast}(s) + s\langle \theta_2, \mu_2 \rangle), \quad \forall s \geq 0,\ \forall \theta_2 \in B(\mathcal{X}_2) \tag{29}$$

As by Lemma 2 $(\alpha_1 \square \alpha_2)^{\circledast} = \alpha_1^{\circledast} + \alpha_2^{\circledast}$, thanks to Corollary 1 again, all we have to prove is

$$\int_{\mathcal{X}_1 \times \mathcal{X}_2} e^{sQ\varphi}\, \mathrm{d}(\mu_1 \otimes \mu_2) = \exp(\alpha_1^{\circledast} + \alpha_2^{\circledast}(s) + s\langle \varphi, \mu_1 \otimes \mu_2 \rangle), \tag{30}$$

for all $s \geq 0$, and $\varphi \in C_b(\mathcal{X}_1 \times \mathcal{X}_2)$.

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

$$\theta_{x_2}(y_1) = Q_2\varphi_{y_1}(x_2) = \inf_{y_2\in Y_2} \{\varphi(y_1,y_2) + c_2(y_2,x_2)\} \tag{31}$$

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

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

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 (31), we have $\sup_{y_1,x_2}|\theta_{x_2}(y_1)| \leq \sup_{y_1}\sup|\varphi_{y_1}| = \sup|\varphi| < \infty$. Therefore, $(x_2,y_1)\mapsto\theta_{x_2}(y_1)$ is an upper semicontinuous bounded function. Consequently, one is allowed to invoke (28) to obtain $\int_{\mathcal{X}_1} e^{sQ_1\theta_{x_2}(x_1)}\,\mu_1(\mathrm{d}x_1) \leq e^{\alpha_1^{\circledast}(s)+s\langle\theta_{x_2},\mu_1\rangle}$ 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 (31).

Remark 2 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,

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

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

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

Inequality (b) is a consequence of (29). This is (30) and concludes the proof of the theorem.

4.2 Product of n spaces

The extension of Theorem 5 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 5 *Let us assume that the cost functions c_i are nonnegative continuous and satisfy* (17). *Suppose that the convex transportation cost inequalities*

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

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

$$\begin{aligned}&\alpha_1\Box\cdots\Box\alpha_n\big(\mathcal{T}_{c_1\oplus\cdots\oplus c_n}(\mu_1\otimes\cdots\otimes\mu_n,\nu)\big)\\ &\quad\le H(\nu\mid\mu_1\otimes\cdots\otimes\mu_n),\quad \forall\nu\in\mathcal{P}(\mathcal{X}_1\times\cdots\times\mathcal{X}_n)\end{aligned}$$

where

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

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

Proof It is a direct consequence of Theorem 5 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.

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 6 *Let us assume that the cost function c is nonnegative continuous and satisfy* (17). *Suppose that the convex transportation cost inequality*

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

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

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

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

Proof This is a direct application of Corollary 5, noting that $\alpha^{\square n}(t) = n\alpha(t/n)$.

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

$$\alpha\left(\mathcal{T}_c(\mu,\nu)\right) \leq H(\nu \mid \mu), \quad \forall \nu \in \mathcal{P}(\mathcal{X}) \tag{32}$$

has the dimension-free tensorization property, if the inequality

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

holds for all $n \in \mathbb{N}^*$.
Clearly, according to Theorem 6, if $\alpha \in \mathcal{C}$ is of the form $\alpha(t) = at$ with $a \geq 0$, then (32) has the dimension-free tensorization property.

Remark 3 Thanks to the same theorem, a seemingly weaker sufficient condition on α for (32) to be dimension-free is $\alpha(t) \leq \inf_{n\geq 1} n\alpha(t/n)$, $t \geq 0$. 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 $t \geq 0$. Therefore, the condition $\alpha(t) \leq \inf_{n\geq 1} n\alpha(t/n)$, $t \geq 0$ is equivalent to $\alpha(t) \leq \alpha'(0)t$, $t \geq 0$. But since α is convex, the converse inequality also holds, that is $\alpha(t) \geq \alpha'(0)t$, $t \geq 0$. Consequently α is of the form $\alpha(t) = at$ with $a \geq 0$.

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

$$\alpha\left(\mathcal{T}_d(\mu, \nu)\right) \le H(\nu \mid \mu), \quad \forall \nu \in \mathcal{P}(\mathcal{X}), \tag{33}$$

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 (33) has the dimension free tensorization property. If μ is a Dirac mass, it is easy to see that (33) 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 (33) has the dimension-free tensorization property, with $\alpha \neq 0$ and let us prove that μ is a Dirac mass. According to Theorem 4, the following inequality

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

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 $\mathrm{dom}\, \alpha^{\circledast}$ (the effective domain of $\alpha^{\circledast}$). But by Jensen inequality, one obtains immediately $\Lambda_\varphi \ge 0$. Thus $\Lambda_\varphi \equiv 0$ on $\mathrm{dom}\, \alpha^{\circledast}$. As $\alpha \neq 0$, $[0, a[\subset \mathrm{dom}\, \alpha^{\circledast}$, 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 $\mu = \delta_{x_0}$. This completes the proof of the proposition.

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

$$\alpha = \max(\alpha_0, \alpha_1). \tag{34}$$

This allows us to separate our investigation into two parts: obtaining α_0 and α_1 which control respectively the small (neighborhood 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 (9).

Definition 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* (8) *holds with* α.

Definition 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 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

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

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

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

Proposition 4 *Under the assumptions of Theorem 2, any deviation function* α *in the class* $\mathcal{C}$ *is a transportation function.*

Proof Let $\alpha \in \mathcal{C}$ be a deviation function. Since $\mathcal{T}(L_n) \ge T_n^\phi$ for all $\phi \in \Phi$, we clearly have $\mathbb{P}(\mathcal{T}(L_n) \ge t) \ge \mathbb{P}(T_n^\phi \ge t)$ for all $t \ge 0$ and n. Therefore, for all ϕ, n and t, $\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(T_n^\phi \ge t) \le \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\mathcal{T}(L_n) \ge t) \le -\alpha(t)$. This implies the statement (d) of Theorem 2, which in turn is equivalent to the statement (a) of Theorem 2, 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 $\mathcal{C}$ and are supposed to be 0 on $(-\infty, 0)$.

Proposition 5 *The first statement is concerned with convex TCIs and the second one with convex $\mathcal{T}$-inequalities.*

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

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

is a transportation function.

(b) *Let us suppose that $\alpha \in \mathcal{C}$ is a transportation function which is extended by $\alpha(t) = 0$ for all $t < 0$, then for all $(\psi, \varphi) \in \Phi$*

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

Remarks.

- In (a), because of Jensen's inequality, one can take $A \ge \int_{\mathcal{X}^2} \exp\beta(c(x,y))\,\mu(\mathrm{d}x)\mu(\mathrm{d}y)$
- About (a), if $c = d \le D < \infty$ is a lower semicontinuous bounded metric, one recovers that

$$\alpha(t) = \begin{cases} 0, & \text{if } t \le D \\ +\infty, & \text{if } t > D \end{cases}$$

  is a transportation function, which is obvious.
- About (b) in the case of a TCI, let us note that $\sup_{(\psi,\varphi)\in\Phi_c}(\varphi(x) + \langle\psi,\mu\rangle) \le \int_{\mathcal{X}} \sup_{\Phi}(\varphi(x) + \psi(y))\,\mu(\mathrm{d}y) \le \int_{\mathcal{X}} c(x,y)\,\mu(\mathrm{d}y)$ for all x. It follows that $\int_{\mathcal{X}} \exp\left[\delta\alpha\left((\varphi(x) + \langle\psi,\mu\rangle)\right)\right]\,\mu(\mathrm{d}x) \le \int_{\mathcal{X}} \exp\left[\delta\alpha\left(\int_{\mathcal{X}} c(x,y)\,\mu(\mathrm{d}y)\right)\right]\,\mu(\mathrm{d}x)$ 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 7.

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

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

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,

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

which with Proposition 4 leads to the desired result.

Let us prove (b). As $\alpha \in \mathcal{C}$ is a transportation function, by Theorem 2 (keeping the notations of Theorem 2) we have

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

By Lemma 3 below, as Λ^*_ϕ is the Cramér transform of $\varphi(X) + \langle \psi, \mu\rangle$ we get

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

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

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

This completes the proof of the proposition.

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

Lemma 3 *Let Z be a real random variable such that $\mathbb{E}e^{\lambda_o|Z|} < \infty$ for some $\lambda_o > 0$. Let $h(z) = \sup_{\lambda\in\mathbb{R}}\{\lambda z - \log \mathbb{E}e^{\lambda Z}\}$ 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 ([9], Lemma 5.1.14). For a proof of the improvement with $(1+\delta)/(1-\delta)$ see [12].

Corollary 6 *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 $\int_{\mathcal{X}} \exp[\gamma \circ \chi(x)]\, \mu(\mathrm{d}x) \leq B < \infty$, 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*

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

If $\alpha \in \mathcal{C}$ *is a transportation function for* c *which is extended by* $\alpha(t) = 0$ *for all* $t < 0$, *then*

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

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

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. Let β be in the class $\mathcal{C}$. We have for all $x_o \in \mathcal{X}$

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

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

$$t \mapsto \max(0, \beta(t) - \log A) = \max(0, 2\gamma(t/2) - 2\log B) \tag{35}$$

is a transportation function for $c = d$.

Now, let us prove (b). Thanks to Kantorovich–Rubinstein equality (7) one can take $\Phi = \{(-\varphi, \varphi); \|\varphi\|_{\mathrm{Lip}} \le 1, \varphi \text{ bounded}\}$. Because of Proposition 5-(b), we have for all bounded φ with $\|\varphi\|_{\mathrm{Lip}} \le 1$:

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

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 $k \geq 0$, with $m := \int_{\mathcal{X}} d(x_o, y)\, \mu(\mathrm{d}y)$, we have

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

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

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

As

$$2\delta\alpha(d(x_o, x)/2) = 2\delta\alpha\left(\frac{d(x_o, x) - m}{2} + \frac{m}{2}\right) \leq \delta[\alpha(d(x_o, x) - m) + \alpha(m)],$$

one sees that

$$\begin{aligned}
&\int_{\mathcal{X}} \exp[2\delta\alpha(d(x_o, x)/2)]\, \mu(\mathrm{d}x) \\
&\quad \leq e^{\delta\alpha(m)} \int_{\mathcal{X}} \exp[\delta\alpha(d(x_o, x) - m)]\, \mu(\mathrm{d}x) \\
&\quad \leq e^{\delta\alpha(m)} (1+\delta)/(1-\delta)
\end{aligned}$$

which leads to

$$\int_{\mathcal{X}} \exp[2\delta\alpha(d(x_o, x)/2)]\, \mu(\mathrm{d}x) < \infty \tag{36}$$

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

$$d_\chi(x, y) = \mathbf{1}_{x \neq y}(\chi(x) + \chi(y)). \tag{37}$$

Remark 5 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 $p \geq 1$, one takes $\chi(x) = 2^{p-1} d(x_o, x)^p$ (see the proof of Corollary 7 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

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

Therefore, any transportation function for d_χ is a transportation function for c. This easy but powerful trick is borrowed from the monograph by Villani ([23], Proposition 7.10).

It has been proved at (35) that if $\int_{\mathcal{X}} \exp[\beta(d_\chi(x_o, x)]\,\mu(dx) \leq C < \infty$ 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

$$\beta(d_\chi(x_o, x)) \leq \gamma \circ \chi(x_o) + \gamma \circ \chi(x) \tag{38}$$

so that $\int_{\mathcal{X}} \exp[\beta(d_\chi(x_o, x)]\,\mu(dx) \leq e^{\gamma\circ\chi(x_o)} B = C$. 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 $\theta(\mathcal{T}_d) \leq \mathcal{T}_c$. 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 (36) with $\alpha \circ \theta$ instead of α completes the proof of the corollary.

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

Corollary 7 ($c = d^p$) *In this statement $c = d^p$ where d is a lower semicontinuous metric and $p \geq 1$.*

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

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

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 $0 \leq u < 2$.

Proof This is Corollary 6 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 (38). One can write $\beta(d_\chi(x_o, x)) \leq \gamma \circ \chi(x_o) + \gamma \circ \chi(x) = \gamma \circ \chi(x)$ 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 6. This completes the proof of the corollary.

Remark 6 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 [22]). As a consequence of Corollary 7-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 6 *Let c be a cost function satisfying* (17) *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*

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

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 (26).

Proof Because of our assumptions, we have $\mathcal{T}_c \le \mathcal{T}_{d_\chi}$, see (37). Hence, it is enough to show that α_o is a transportation function for d_χ. But this follows from Lemma 4 below and Corollary 3.

The last statement follows from a simple manipulation on the definition of the Orlicz norm $\|\chi\|_\rho$. This completes the proof of the proposition.

The following lemma has been used in the previous proof.

Lemma 4 *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 (7), 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 [12]) that this is equivalent to $|\varphi(x) - a| \le \chi(x), \forall x$ for some real a. Therefore,

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

which is the desired result.

5.4 An application: T_1-inequalities

A T_1-inequality is a TCI with $c = d$. Let us denote $\mathcal{P}_d(\mathcal{X}) = \{\nu \in \mathcal{P}(\mathcal{X}); \int_{\mathcal{X}} d(x_*, x)\, \nu(\mathrm{d}x) < \infty$ 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

$$\alpha(\mathcal{T}_d(\mu,\nu)) \le H(\nu \mid \mu), \quad \forall \nu \in \mathcal{P}_d(\mathcal{X}). \tag{39}$$

Theorem 7 (T_1-inequalities) *Let d be a lower semicontinuous metric. Suppose that $a \ge 0$ satisfies $\int_{\mathcal{X}} e^{ad(x_o,x)}\, \mu(\mathrm{d}x) \le 2$ for some $x_o \in \mathcal{X}$ and that $\gamma \in \mathcal{C}$ satisfies $\int_{\mathcal{X}} e^{\gamma(d(x_1,x))}\, \mu(\mathrm{d}x) \le B < \infty$ for some $x_1 \in \mathcal{X}$, then*

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

satisfies (39).

Conversely, if a function α in the class $\mathcal{C}$ satisfies (39), *then*

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

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

Proof Gathering Corollary 7-a, Proposition 6 and the trick (34) gives us the first statement. The second statement is a particular instance of Corollary 7-b. This completes the proof of the theorem.

Note that by Proposition 2 we know that it is impossible that α escapes from a quadratic growth at the origin.

Theorem 7 extends the integral criteria for the usual $T_1(C)$-inequality in [10] and [5]. Nevertheless, the control of the constant C is handled more carefully in these cited papers.

In a forthcoming paper (see the PhD manuscript [12]), one of the author has obtained the following result which is very much in the spirit of [10] and [5].

Theorem 8 *Suppose that* $c(x,y) = d^p(x,y)$, *that* α *satisfies* (25) *for some* $a > 0$ *and that* $\sup\{\alpha^{\circledast}(t); t : \alpha^{\circledast}(t) < +\infty\} = +\infty$. *Then, the following statements are equivalent*:

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

Further details concerning the relation between b_1 and b_2 can be found in [12].

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

In this section, we give some applications of transportation-cost inequalities. The first application, Theorem 9 is an easy extension of a well known result of Marton. The second one, Theorem 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

Theorems 9 and 10 both rely on the following elementary lemma.

Lemma 5 *Let* $\mu \in \mathcal{P}(\mathcal{X})$ *be such that* $\int_{\mathcal{X}} d(x_o,x)\,\mu(\mathrm{d}x) < +\infty$, *for some* (*and thus all*) $x_o \in \mathcal{X}$, *and suppose that the* T_1*- inequality*

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

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

$$\mu\,(\varphi \ge \langle \varphi, \mu\rangle + t) \le e^{-\alpha(t)}, \quad \forall t > 0. \tag{40}$$

Proof Let φ a 1-Lipschitz function. For every $n \ge 1$, let us consider $\varphi_n = \varphi \vee n \wedge -n$. According to Theorem 4, one has

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

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

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

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

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

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 $r_o \geq 0$ such that

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

or equivalently

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

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 Marton, see [16]. Her result extends as follows.

Theorem 9 *Let $\mu \in \mathcal{P}(\mathcal{X})$ be such that $\int_{\mathcal{X}} d(x_o, x)\, \mu(\mathrm{d}x) < +\infty$ for some $x_o \in \mathcal{X}$, 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 $\mu(A) \in (0, 1)$, one has the following concentration of measure inequality:

$$\mu(A^r) \geq 1 - e^{-\alpha(r - r_A)}, \quad \forall r \geq r_A, \tag{41}$$

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

Remark 7 According to the assumptions made on α, one sees that the function α^{-1} is well defined on $(0, +\infty)$. The number r_A is thus well defined too.

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 [12].

Proof The function $x \mapsto d(x,A)$ is 1-Lipschitz. Thus, according to Lemma 5,

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

In order to derive (41), 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

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

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 $\mathrm{d}\mu_A = \frac{\mathbf{1}_A}{\mu(A)}\mathrm{d}\mu$; clearly $\mu_A(A) = 1$, so

$$\inf\left\{H(\nu \mid \mu) : \nu(A) = 1\right\} \leq H(\mu_A \mid \mu). \tag{42}$$

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

Note that $d(\cdot,A)$ is unbounded so that the inequality $\int_{\mathcal{X}} d(\cdot,A)\,\mathrm{d}\mu - \int_{\mathcal{X}} d(\cdot,A)\,\mathrm{d}\nu \leq \mathcal{T}_d(\mu,\nu)$ needs to be justified. Let π be a probability on $\mathcal{X}^2$ with marginals μ and ν, then $\int_{\mathcal{X}} d(\cdot,A)\,\mathrm{d}\mu - \int_{\mathcal{X}} d(\cdot,A)\,\mathrm{d}\nu = \iint_{\mathcal{X}^2} d(x,A) - d(y,A)\,\pi(\mathrm{d}x\mathrm{d}y) \leq \iint_{\mathcal{X}^2} d(x,y)\,\pi(\mathrm{d}x\mathrm{d}y)$. 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 (42) holds with equality:

$$\inf\left\{H(\nu \mid \mu) : \nu(A) = 1\right\} = H(\mu_A \mid \mu). \tag{43}$$

In other words, μ_A is Csiszár's I-projection of μ on $\{\nu \in \mathcal{P}(\mathcal{X}) : \nu(A) = 1\}$, see [7,8].

If ν is such that $\nu(A) = 1$, one has

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

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 (43).

6.3 TCIs and deviations bounds for empirical processes.

Lemma 6 *Let $p \geq 1$ and $\mu \in \mathcal{P}(\mathcal{X})$ be such that $\int_{\mathcal{X}} d^p(x_o, x)\, \mu(\mathrm{d}x) < +\infty$, for some $x_o \in \mathcal{X}$, and suppose that the following inequality*

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

holds. Then for all function $Z : \mathcal{X}^n \to \mathbb{R}$ which is $n^{-1/p}$-Lipschitz with respect to the metric $(x, y) \mapsto \sqrt[p]{\sum_{i=1}^n d^p(x_i, y_i)}$, one has

$$\mu^{\otimes n}\left(Z \geq \langle \mu^{\otimes n}, Z\rangle + t\right) \leq e^{-n\alpha(t^p)}, \quad \forall t \geq 0 \tag{44}$$

Proof According to the tensorization property stated in Theorem 6, $\mu^{\otimes n}$ satisfies the TCI

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

where $d^{p\oplus n}(x, y) = \sum_{i=1}^n d(x_i, y_i)^p$, for all $x, y \in \mathcal{X}^n$. Applying Jensen's inequality, one gets immediately : $\mathcal{T}_{d^{p\oplus n}}(\mu^{\otimes n}, \nu) \geq \mathcal{T}_{d_{p,n}}(\mu^{\otimes n}, \nu)^p, \forall \nu \in \mathcal{P}(X^n)$, where $d_{p,n}$ is the metric defined by $d_{p,n}(x, y) = \sqrt[p]{\sum_{i=1}^n d^p(x_i, y_i)}$. Thus $\mu^{\otimes n}$ satisfies the following T_1-inequality:

$$\widetilde{\alpha}\left(\mathcal{T}_{d_{p,n}}(\mu^{\otimes n}, \nu)\right) \leq H(\nu \mid \mu^{\otimes n}), \quad \forall \nu \in \mathcal{P}(\mathcal{X}^n),$$

where $\widetilde{\alpha} \in \mathcal{C}$ is defined by $\widetilde{\alpha}(t) = n\alpha\left(\frac{t^p}{n}\right)$. The function $n^{1/p}Z$ being 1-Lipschitz with respect to $d_{p,n}$, it follows from Lemma 5 that

$$\mu^{\otimes n}\left(n^{1/p}Z \geq n^{1/p}\langle \mu^{\otimes n}, Z\rangle + n^{1/p}t\right) \leq e^{-\widetilde{\alpha}(n^{1/p}t)} = e^{-n\alpha(t^p)},$$

for all $t \geq 0$, which completes the proof.

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

$$Z_n^{\mathcal{G}} := \sup_{\varphi \in \mathcal{G}} \left\{ \left| \frac{1}{n} \sum_{i=1}^n \varphi(X_i) - \int_{\mathcal{X}} \varphi \, \mathrm{d}\mu \right| \right\}. \tag{45}$$

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

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

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, one immediately obtains the following theorem.

Theorem 10 *Let $\mu \in \mathcal{P}(\mathcal{X})$ be such that $\int_{\mathcal{X}} d(x_o, x)\, \mu(\mathrm{d}x) < +\infty$, for some $x_o \in \mathcal{X}$, 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* (45) *satisfies the following inequality*

$$\mathbb{P}\left(Z_n^{\mathcal{G}} \geq \mathbb{E}\left[Z_n^{\mathcal{G}}\right] + t\right) \leq e^{-n\alpha(t)}, \quad \forall t \geq 0. \tag{46}$$

The literature about the deviations of empirical processes is huge. For a good overview of this subject, one can read Massart's Saint-Flour lecture notes [18].

Now, if $(\mathcal{X}, \|\cdot\|)$ is a separable Banach space, and $\mu \in \mathcal{P}(\mathcal{X})$ such that $\int_{\mathcal{X}} \|x\|\, \mathrm{d}\mu < +\infty$ 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

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

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

Theorem 11 *Let $\mu \in \mathcal{P}(\mathcal{X})$ be such that $\int_{\mathcal{X}} \|x\|\, \mu(\mathrm{d}x) < +\infty$, and suppose that the T_1-inequality*

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

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

$$\mathbb{P}\left(Z_n \geq \mathbb{E}\left[Z_n\right] + t\right) \leq e^{-n\alpha(t)}, \quad \forall t \geq 0. \tag{47}$$

The preceding tools can also be used to derive quantitative versions of Sanov's theorem involving Wasserstein's metrics. Recall that, if $p \geq 1$, then

$W_p(\nu_1,\nu_2) := \sqrt[p]{\mathcal{T}_{d^p}(\nu_1,\nu_2)}$ defines a metric on $\mathcal{P}_p(\mathcal{X})$, the set of probability measures on $\mathcal{X}$ which integrate $d^p(x_o,\,.\,)$ for some (and thus all) x_o (see Chap. 7 of [23]).

Theorem 12 *Let $\mu \in \mathcal{P}(\mathcal{X})$ be such that $\int_{\mathcal{X}} d^p(x_o,x)\,\mu(\mathrm{d}x) < +\infty$, for some $x_o \in \mathcal{X}$, and suppose that the following inequality*

$$\alpha\left(\mathcal{T}_{d^p}(\mu,\nu)\right) \le H(\nu \mid \mu), \quad \forall \nu \in \mathcal{P}(\mathcal{X}),$$

holds for some $p \ge 1$ and $\alpha \in \mathcal{C}$. Then

$$\mathbb{P}\left(W_p(L_n,\mu) \ge \mathbb{E}\left[W_p(L_n,\mu)\right] + w\right) \le e^{-n\alpha(w^p)}, \quad \forall w \ge 0.$$

Proof Let us denote $L_n^x = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, for all $x \in \mathcal{X}^n$. According to Lemma 6, it suffices to show that $x \mapsto W_p(L_n^x,\mu)$ is $n^{-1/p}$-Lipschitz with respect to the metric $d_{p,n}(x,y) = \sqrt[p]{\sum_{i=1}^n d^p(x_i,y_i)}$. Take $x,y \in \mathcal{X}^n$; according to the triangle inequality, $\left|W_p(L_n^x,\mu) - W_p(L_n^y,\mu)\right| \le W_p(L_n^x,L_n^y)$. As $\mathcal{T}_{d^p}(\mu,\nu)$ is jointly convex, we have

$$W_p(L_n^x,L_n^y) \le \sqrt[p]{\frac{1}{n}\sum_{i=1}^n \mathcal{T}_{d^p}(\delta_{x_i},\delta_{y_i})} = n^{-1/p}\sqrt[p]{\sum_{i=1}^n d^p(x_i,y_i)},$$

which completes the proof.

Remark 8 In order to obtain precise deviations results for $Z_n^{\mathcal{G}}$ (resp. Z_n), one must be able to estimate the terms $\mathbb{E}\left[Z_n^{\mathcal{G}}\right]$, $\mathbb{E}\left[Z_n\right]$ or $\mathbb{E}\left[W_p(L_n,\mu)\right]$.

Let us give some examples.

Example 2 (Quantitative versions of Sanov theorem) In this example, $\mathbb{R}^q$ will be furnished with a norm $\|\cdot\|$. The metric associated to this norm will be denoted by $d(\,.\,,\,.\,)$. The following theorem is Theorem 10.2.1 of [20] (volume II).

Theorem 13 *Let μ be a probability measure on $\mathbb{R}^q$ such that*

$$c := \int \|x\|^{q+5}\,\mathrm{d}\mu < +\infty. \tag{48}$$

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

$$\mathbb{E}\left[\mathcal{T}_{d^2}(L_n,\mu)\right] \le D n^{-\frac{2}{q+4}}. \tag{49}$$

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

Corollary 8 *Let μ be a probability on $(\mathbb{R}^q, \|\cdot\|)$, satisfying (48) and the inequality*

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

for some $p \in [1,2]$ and some $\alpha \in \mathcal{C}$. Then, the following inequality holds:

$$\mathbb{P}\left(W_p(L_n,\mu) \geq w\right) \leq \exp\left[-n\alpha\left(\left(w - D/n^{\frac{1}{q+4}}\right)^p\right)\right]$$

for all $w > 0$ and $n \geq (D/w)^{q+4}$ where D is the constant of (49).

Proof Noting that $\mathbb{E}\left[W_p(L_n,\mu)\right] \leq \mathbb{E}[W_2(L_n,\mu)] \leq \sqrt{\mathbb{E}\left[\mathcal{T}_{d^2}(L_n,\mu)\right]}$, for all $p \in [1,2]$, the result follows immediately from Theorems 12 and 13.

In [4], Bolley et al. have shown that if μ satisfies the inequality $T_p(C)$ (see (1)) with $1 \leq p \leq 2$, then there are $n_o > 0$ and $\widetilde{C} > 0$ such that the inequality

$$\mathbb{P}(W_p(L_n,\mu) \geq w) \leq e^{-n\widetilde{C}w^2}$$

holds for $n \geq n_o w^{-(q+2)}$. Corollary 8 is quite similar but its proof is completely different.

Example 3 (Deviations bounds for empirical means) Let $\mathcal{X}$ be a separable Banach space and consider

$$Z_n = \left\| \frac{1}{n}\sum_{i=1}^n X_i - \int_{\mathcal{X}} x \,\mathrm{d}\mu \right\|, \tag{50}$$

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

$$\mathbb{E}\left[\|Y_1 + \cdots + Y_n\|^p\right] \leq b\left[\mathbb{E}\left[\|Y_1\|^p\right] + \cdots + \mathbb{E}\left[\|Y_n\|^p\right]\right]. \tag{51}$$

If $\mathcal{X}$ is of type p and $\mathbb{E}\left[\|X_1\|^p\right] < +\infty$, then one can deduce immediately from (51) the following control:

$$\mathbb{E}\left[Z_n\right] \leq \frac{1}{n^{1-1/p}}\left(b\mathbb{E}\left[\|X_1 - \mathbb{E}[X_1]\|^p\right]\right)^{1/p}. \tag{52}$$

Controls like (52) can be used in Theorem 11 to derive precise deviations bounds for empirical means. Let us conclude this section with a concrete example.

Theorem 14 *Let μ be a probability measure on a separable Banach space $(\mathcal{X}, \|\cdot\|)$ such that $\int_{\mathcal{X}} e^{a\|x\|}\,\mu(\mathrm{d}x) < +\infty$, for some $a > 0$. Then, for all sequence X_i of iid random variables with law μ, one has*

$$\mathbb{P}\left(Z_n \geq \mathbb{E}[Z_n] + t\right) \leq e^{-n\left(\sqrt{1+\frac{t}{M}}-1\right)^2}, \quad \forall t > 0, \tag{53}$$

where Z_n is defined by (50) *and* $M := \inf\left\{b > 0 : \iint_{\mathcal{X}^2} e^{\frac{\|x-y\|}{b}}\,\mu(\mathrm{d}x)\mu(\mathrm{d}y) \leq 2\right\}$.

Proof According to Corollary 4, 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 11, the result follows immediately.

Inequality (53) is very close to a well known inequality by Yurinskii ([24], Theorem 2.1). Under the same assumptions on μ, one can easily derive from Yurinskii's result the following bound :

$$\mathbb{P}\left(Z_n \geq \mathbb{E}\left[Z_n\right] + t\right) \leq \exp\left(-\frac{1}{8}\frac{nt^2}{2M_o^2 + tM_o}\right), \quad \forall t > 0, \tag{54}$$

where $M_o = \inf\left\{b > 0 : \int_{\mathcal{X}} e^{\frac{\|x\|}{b}}\,\mu(\mathrm{d}x) \leq 2\right\}$. To compare (53) and (54) first note that

$$\left(\sqrt{1+u}-1\right)^2 \geq \frac{u^2}{2(2+u)}, \quad \forall u > 0, \tag{55}$$

(this is left to the reader). Next, let us show that $M \leq 2M_o$. This follows from the following inequality:

$$\iint\limits_{\mathcal{X}^2} e^{\frac{\|x-y\|}{2M_o}}\,\mu(\mathrm{d}x)\mu(\mathrm{d}y) \overset{(i)}{\leq} \left(\int\limits_{\mathcal{X}} e^{\frac{\|x\|}{2M_o}}\,\mu(\mathrm{d}x)\right)^2 \overset{(ii)}{\leq} \int\limits_{\mathcal{X}} e^{\frac{\|x\|}{M_o}}\,\mu(\mathrm{d}x) \overset{(iii)}{\leq} 2,$$

where (i) comes from the triangle inequality, (ii) from Jensen inequality and (iii) from the definition of M_o. Thanks to (55), one obtains

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

Thus, (53) is a little bit stronger than (54).

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 Sect. 5. See in particular Remark 4.

7.1 A deviation function is a transportation function

In this section, we give a rigorous proof at Theorem 15 of the Recipe 2 for an increasing deviation function which may possibly be not convex. This extends Proposition 4.

Theorem 15 *Let us assume* (14) *and* (15).

(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 from 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 $t \geq 0$,

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

where the last equality holds because α is increasing. This means that α is a deviation function. This completes the proof of the theorem.

Remark 9 Note that we didn't use the specific form (9) 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 $\{L_n\}$.

Let us specialize Theorem 15 to the case where $c = d^p$. Let d be a metric on $\mathcal{X}$ which turns it into a Polish space and $1 \leq p < \infty$. We denote $\mathcal{N}_{d^p}$ the set of all probability measures ν on $\mathcal{X}$ such that $\int_{\mathcal{X}} d(x_o,x)^p\,\nu(\mathrm{d}x) < \infty$ for some x_o in $\mathcal{X}$.

Corollary 9 **($c = d^p$)** *Let μ in $\mathcal{P}(\mathcal{X})$ satisfy $\int_{\mathcal{X}} e^{ad(x_o,x)^p}\,\mu(\mathrm{d}x) < \infty$ for all $a > 0$ and some x_o in $\mathcal{X}$. Then, a left continuous increasing function $\alpha : [0,\infty) \to [0,\infty]$ satisfies $\alpha(\mathcal{T}_{d^p}(\mu,\nu)) \le H(\nu|\mu)$ for all ν in $\mathcal{N}_{d^p}$ if and only if $\alpha(t) \le -\limsup_{n\to\infty} \frac{1}{n}\log\mathbb{P}(\mathcal{T}_{d^p}(\mu,L_n) \ge t)$ for all $t \ge 0$.*

In other words, the left continuous version of $-\limsup_{n\to\infty} \frac{1}{n}\log\mathbb{P}(\mathcal{T}_{d^p}(\mu,L_n) \ge t)$ is the best transportation function.

Proof It is known that 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 $|\varphi(x)| \le c(1+d(x_o,x)^p), \forall x$ for some constant c, see the annex of Bolley's PhD Thesis manuscript [3]. Therefore, $\nu \mapsto \mathcal{T}_{d^p}(\mu,\nu)$ is continuous on $\mathcal{P}_{\mathcal{F}}$ and one can apply Theorem 15. This completes the proof of the corollary.

In the case where μ satisfies $T_2(C)$, see (1), with Corollary 8 one sees that $\limsup_{n\to\infty} \frac{1}{n}\log\mathbb{P}(\mathcal{T}_{d^2}(\mu,L_n) \ge t) \le -Dt$ for all $t \ge 0$ and some positive D. In particular, this holds when μ is the Gaussian measure. Corollary 9 states the converse result provided that $\int_{\mathcal{X}} e^{ad(x_o,x)^2}\,\mu(\mathrm{d}x) < \infty$ for all $a > 0$, which rules the Gaussian measure out.

7.2 The transportation function J_Φ

With Theorem 15 in hand, it is enough to compute a deviation function α to obtain the TCI

$$\alpha(\mathcal{T}(\nu)) \le H(\nu \mid \mu), \quad \forall \nu \in \mathcal{P}_{\mathcal{F}} \tag{56}$$

But these functions may be rather hard to compute because of the sup in the definition (9) of

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

However, it is shown at Theorem 16 below, that more can be said about transportation functions.

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

(i) We assume (14):
$$\int_{\mathcal{X}} e^{\varphi}\,\mathrm{d}\mu < \infty, \quad \forall\varphi \in \mathcal{F}.$$

(ii) We assume (15):
$$(0,0) \in \Phi \subset \mathcal{F}\times\mathcal{F},$$

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

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

For all $(\psi, \varphi) \in \Phi$, let us define

$$\Lambda(\varphi) := \log \int_{\mathcal{X}} e^{\varphi} \, \mathrm{d}\mu, \quad \Lambda_{\psi,\varphi}(s) := \Lambda(s\varphi) + s\langle \psi, \mu \rangle, \; s \in \mathbb{R},$$

and

$$J_{\psi,\varphi}(t) = \sup_{s \in \mathbb{R}} \{st - \Lambda_{\psi,\varphi}(s)\}, \quad t \in \mathbb{R}.$$

Remark 10 As a consequence of Assumptions (A), $\{L_n\}$ obeys the LDP in $\mathcal{P}_{\mathcal{F}}$ with the rate function $H(\nu \mid \mu) = \Lambda^*(\nu) = \sup_{\varphi \in \mathcal{F}} \{\langle \varphi, \nu \rangle - \Lambda(\varphi)\}, \nu \in \mathcal{P}_{\mathcal{F}}$ and thanks to Cramer theorem, $J_{\psi,\varphi}$ is the LD rate function of $\{\langle \varphi, L_n \rangle + \langle \psi, \mu \rangle\}_{n \geq 1}$.

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

$$\begin{aligned} J_\Phi(t) &:= \widetilde{J}_\Phi(t^-), t > 0 \quad \text{where} \\ \widetilde{J}_\Phi(t) &:= \inf_{(\psi,\varphi) \in \Phi} J_{\psi,\varphi}(t) \in [0, \infty], t \geq 0 \end{aligned} \tag{57}$$

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 4 (of the functions J_Φ, J_{tr} and J_{dev}).

- 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_{tr} *is the* best transportation function. *Clearly, it is the left continuous version of the increasing function*

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

- J_{dev} *is the* best deviation function. *Clearly, it is the left continuous version of the increasing function*

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

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

Theorem 16 *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 17-a and Corollary 10-a,b which will be proved below.

Theorem 17 *Suppose that Assumptions* (A) *hold.*

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

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

(b) *If in addition $\mathcal{T}$ is continuous on $\mathcal{P}_\mathcal{F}$, then $J_\Phi = J_{\text{tr}} = J_{\text{dev}}$.*

Proof (a) As $\nu \mapsto \langle \varphi, \nu\rangle + \langle \psi, \mu\rangle$ is continuous, it follows from Remark 10 and 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 (9) 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) Theorem 15-(b) states that $J_{\text{tr}} = J_{\text{dev}}$. Hence, it is enough to prove that $J_\Phi = J_{\text{dev}}$. Because of part (a) of the present theorem, J_Φ is a transportation function, and by part (b) of Theorem 15, it is also a deviation function. Therefore, $J_\Phi \le J_{\text{dev}}$ and it remains to prove that $J_{\text{dev}} \le J_\Phi$. By the LD lower bound for $\{\langle \varphi, L_n\rangle + \langle \psi, \mu\rangle\}$, for all $t \ge 0$,

$$\begin{aligned} -\inf_{r>t} J_{\psi,\varphi}(r) &\le \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\langle \varphi, L_n\rangle + \langle \psi, \mu\rangle > t) \\ &\le \limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\sup_{(\psi,\varphi)\in\Phi} \langle \varphi, L_n\rangle + \langle \psi, \mu\rangle \ge t \right) \\ &\le -J_{\text{dev}}(t). \end{aligned}$$

Thus $J_{\text{dev}}(t) \le \inf_{r>t} J_{\psi,\varphi}(r)$, for all $t \ge 0$ and consequently

$$J_{\text{dev}}(t) \le \inf_{(\psi,\varphi)\in\Phi} \inf_{r>t} J_{\psi,\varphi}(u) = \inf_{r>t} \inf_{(\psi,\varphi)\in\Phi} J_{\psi,\varphi}(u) = \widetilde{J}_\Phi(t^+).$$

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

7.3 Connections with Theorem 2

Let us first give an alternative proof of criterion $(b) \Rightarrow (a)$ of Theorem 2.

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

$$\Lambda_\Phi(s) = \sup_{(\psi,\varphi)\in\Phi} \Lambda_{\psi,\varphi}(s) \quad \text{with } \Lambda_{\psi,\varphi}(s) = \Lambda(s\varphi) + s\langle\psi,\mu\rangle, \;\; s \geq 0 \tag{58}$$

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

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

is also in $\mathcal{C}$. Thanks to formula (57), 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

$$\Lambda_\Phi^\circledast \leq J_\Phi. \tag{59}$$

As J_Φ is a transportation function (Theorem 17), so is $\Lambda_\Phi^\circledast$.

The criterion $(b) \Rightarrow (a)$ of Theorem 2 follows from the above considerations. Indeed, (b) states that $\Lambda_\Phi \leq \alpha^\circledast$. Therefore, with (59): $\alpha \leq \Lambda_\Phi^\circledast \leq J_\Phi$. Hence, α is a transportation function.

An easy consequence of Theorem 2 is the following

Corollary 10 *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 2 is $\alpha^\circledast = \Lambda_\Phi$, see (58). Because of the equivalence $(a) \Leftrightarrow (b)$ of Theorem 2, 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

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

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 17-(b).

References

1. Bobkov, S., Götze, F.: Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. J. Funct. Anal. **163**, 1–28 (1999)
2. Bobkov, S.G., Gentil, I., Ledoux, M.: Hypercontractivity of Hamilton-Jacobi equations. J. Math. Pures Appl. **80**(7), 669–696 (2001)
3. Bolley, F.: Applications du transport optimal à des problèmes de limites de champ moyen (2005). PhD Thesis, École Normale Supérieure de Lyon
4. Bolley, F., Guillin, A., Villani, C.: Quantitative concentration inequalities for empirical measures on non-compact spaces (2005). Preprint. Available online via http://www.latp.univ-mrs.fr/~guillin/index3.html. To appear in Prob. Theory Related Fields
5. Bolley, F., Villani, C.: Weighted Csiszár-Kullback-Pinsker inequalities and applications to transportation inequalities. Ann. Faculté Sci. Toulouse. **14**, 331–352 (2005)
6. Borell, C.: The Brunn-Minkowski inequality in Gauss space. Invent. Math. **30**, 207–216 (1975)
7. Csiszár, I.: *I*-divergence geometry of probability distributions and minimization problems. Ann. Prob. **3**, 146–158 (1975)
8. Csiszár, I.: Sanov property, generalized *I*-projection and a conditional limit theorem. Ann. Prob. **12**, 768–793 (1984)
9. Dembo, A., Zeitouni, O.: Large deviations techniques and applications, 2nd edn. Applications of Mathematics vol. 38. Springer, Berlin Heidelberg New York (1998)
10. Djellout, H., Guillin, A., Wu, L.: Transportation cost-information inequalities for random dynamical systems and diffusions. Ann. Prob. **32**(3B), 2702–2732 (2004)
11. Eichelsbacher, P., Schmock, U.: Large deviations of U-empirical measures in strong topologies and applications. Ann. Inst. Henri Poincaré **38**(5), 779–797 (2002)
12. Gozlan, N.: Principe conditionnel de Gibbs pour des contraintes fines approchées et inégalités de transport (2005). PhD Thesis, Université de Paris 10
13. Ledoux, M.: The concentration of measure phenomenon. Mathematical Surveys and Monographs, vol. 89. American Mathematical Society, Providence (2001)
14. Léonard, C.: A set of lecture notes on convex optimization with some applications to probability theory (2006). Incomplete draft. Available online via http://www.cmap.polytechnique.fr/~leonard/
15. Léonard, C., Najim, J.: An extension of Sanov's theorem: application to the Gibbs conditioning principle. Bernoulli **8**(6), 721–743 (2002)
16. Marton, K.: A simple proof of the blowing-up lemma. IEEE Trans. Inf. Theory **32**, 445–446 (1986)
17. Marton, K.: Bounding $\bar{d}$-distance by informational divergence: a way to prove measure concentration. Ann. Prob. **24**, 857–866 (1996)

18. Massart, P.: Saint-Flour Lecture Notes (2003). Available online via http://www.math.u-psud.fr/˜massart/
19. Otto, F., Villani, C.: Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. J. Funct. Anal. **173**, 361–400 (2000)
20. Rachev, S., Rüschendorf, L.: Mass transportation problems. vol I: Theory, Vol. II : Applications. Probability and its applications. Springer, Berlin Heidelberg New York (1998)
21. Sanov, I.N.: On the probability of large deviations of random variables. Selected Transl. Math. Statist. Prob. **1**, 214–244 (1961)
22. Talagrand, M.: Transportation cost for gaussian and other product measures. Geom. Funct. Anal. **6**, 587–600 (1996)
23. Villani, C.: Topics in optimal transportation. Graduate Studies in Mathematics, vol. 58. American Mathematical Society, Providence (2003)
24. Yurinskii, V.: Exponential inequalities for sums of random vectors. J. Multivariate Anal. **6**, 473–499 (1976)