



**HAL**  
open science

# KANTOROVICH DUALITY FOR GENERAL TRANSPORT COSTS AND APPLICATIONS

Nathael Gozlan, Cyril Roberto, Paul-Marie Samson, Prasad Tetali

► **To cite this version:**

Nathael Gozlan, Cyril Roberto, Paul-Marie Samson, Prasad Tetali. KANTOROVICH DUALITY FOR GENERAL TRANSPORT COSTS AND APPLICATIONS. 2014. hal-01098114v1

**HAL Id: hal-01098114**

**<https://hal.science/hal-01098114v1>**

Preprint submitted on 22 Dec 2014 (v1), last revised 24 Dec 2015 (v4)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# KANTOROVICH DUALITY FOR GENERAL TRANSPORT COSTS AND APPLICATIONS

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

ABSTRACT. We introduce a general notion of transport cost that encompasses many costs used in the literature (including the classical one and weak transport costs introduced by Talagrand and Marton in the 90's), and prove a Kantorovich type duality theorem. As a by-product we obtain various applications in different directions: we give a short proof of a result by Strassen on the existence of a martingale with given marginals, we characterize the associated transport-entropy inequalities together with the log-Sobolev inequality restricted to convex/concave functions. Some explicit examples of discrete measures satisfying weak transport-entropy inequalities are also given.

## 1. INTRODUCTION

Optimal transport is a very active field with many connections and applications to other areas of mathematics, including Probability Theory, Analysis, Geometry... One possible reference for an introduction and related topics is the book by Villani [37]. A key tool in the whole theory is the celebrated Kantorovich duality theorem which, if  $\omega: X \times X \rightarrow [0, \infty)$  denotes a (say) continuous cost function on a complete and separable metric space  $(X, d)$ , asserts that, given any two probability measures  $\mu, \nu$  on  $X$ ,

$$\inf_{\pi \in \Pi(\mu, \nu)} \left\{ \iint \omega(x, y) \pi(dx dy) \right\} = \sup_{\varphi \in \mathcal{C}_b(X)} \left\{ \int Q_\omega \varphi d\mu - \int \varphi d\nu \right\},$$

where  $\Pi(\mu, \nu)$  denotes the set of probability measures  $\pi$ , on  $X \times X$  (couplings), with first marginal  $\mu$  and second marginal  $\nu$ , and  $\mathcal{C}_b(X)$  is the set of all bounded continuous functions  $\varphi: X \rightarrow \mathbb{R}$ , with

$$Q_\omega \varphi(x) := \inf_{y \in X} \{ \varphi(y) + \omega(x, y) \}, \quad x \in X.$$

Note that the assumptions in the above statement can be weakened (see [37], [28] and the references therein). The associated transport cost  $\mathcal{T}_\omega(\nu, \mu) := \inf_{\pi \in \Pi(\mu, \nu)} \{ \iint \omega(x, y) \pi(dx dy) \}$ , with  $\omega(x, y) = d(x, y)^2$  plays an important role in many applications (again, see *e.g.* [37]). However, in many interesting situations, including the discrete setting, one would like to deal with more general transport costs.

The main aim of the present paper is to give a unified general setting that includes many costs used in the literature and to prove the corresponding duality theorem. As a consequence, we shall give different applications in very different contexts.

---

*Date:* December 22, 2014.

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

*Key words and phrases.* Duality, Transport inequalities, logarithmic-Sobolev inequalities, metric spaces.

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

To be more specific, besides the classical case mentioned above, our setting should include the following weak transport costs introduced respectively by Talagrand [34] and by Marton [22]. In the definitions below, if  $\pi \in \Pi(\mu, \nu)$  is a coupling between  $\mu$  and  $\nu$ , its disintegration with respect to its first marginal is written as follows

$$\pi(dxdy) = p_x(dy) \mu(dx),$$

where  $x \mapsto p_x$  is a measurable probability kernel. If  $\alpha, \beta: \mathbb{R}^+ \rightarrow \mathbb{R}^+$  are convex functions, one defines the following transport-like cost functionals :

$$(1.1) \quad \widehat{\mathcal{T}}_\beta(\nu|\mu) = \inf_{\pi \in \Pi(\mu, \nu)} \iint \beta \left( \mathbf{1}_{x \neq y} \frac{dp_x}{d\mu_0}(dy) \right) \mu_0(dy) \mu(dx),$$

where in addition  $\mu, \nu$  are assumed to be absolutely continuous with respect to a fixed (reference) probability measure  $\mu_0$ ; and

$$(1.2) \quad \widetilde{\mathcal{T}}_\alpha(\nu|\mu) = \inf_{\pi \in \Pi(\mu, \nu)} \iint \alpha \left( \int \mathbf{1}_{x \neq y} p_x(dy) \right) \mu(dx).$$

Also, when  $X = \mathbb{R}^m$  is equipped with some norm  $\|\cdot\|$ , our setting should include the following variant ( $\theta: \mathbb{R}^m \rightarrow [0, \infty]$  is a convex function)

$$(1.3) \quad \overline{\mathcal{T}}_\theta(\nu|\mu) = \inf_{\pi \in \Pi(\mu, \nu)} \int \theta \left( x - \int y p_x(dy) \right) \mu(dx).$$

The transport cost defined in (1.1) is implicitly used by Talagrand to reach concentration results on product spaces (see [34, Theorem 4.2]). As a main application, it provides Bernstein type deviation inequalities for supremum of empirical processes. This method was further developed by the third named author [31, Theorem 1.1] to get transport type inequalities and optimal deviation bounds for supremum of empirical processes (see Section 7.5, where some of these results are recalled).

On the other hand, the transport cost defined in (1.2) was used by many authors (see Marton [21, 23, 22], Dembo [7], Samson [29, 30, 31], Wintemberger [38]) to deal with concentration results, including some applications in discrete settings [32], and as a discrete counterpart of the usual  $\mathcal{T}_2$  Kantorovich-Wasserstein distance in some displacement convexity property of the entropy along interpolating paths, on graphs [15].

Finally, we shall introduce the transport cost  $\overline{\mathcal{T}}$  (the terminology *bar* makes reference to the *barycenter* entering its definition) since, for  $\theta(x) = |x|$ ,  $x \in \mathbb{R}$ , it will appear as a key ingredient in a new two-line proof (which basically reduces only to the corresponding Kantorovich duality theorem) of a result by Strassen [33] on the existence of a martingale with given marginals.

Having these applications in mind, if  $\mathcal{P}(X)$  denotes the set of all probability measures on  $X$ , a cost is a function  $c: X \times \mathcal{P}(X) \rightarrow [0, \infty]$  and its associated transport cost is

$$\mathcal{T}_c(\nu|\mu) := \inf_{\pi \in \Pi(\mu, \nu)} \int c(x, p_x) \mu(dx).$$

Note that, as expected, this definition encompasses the classical case (with  $c(x, p) := \int_X \omega(x, y) p(dy)$ ), the weak transport costs introduced by Talagrand and Marton (with  $c(x, p) := \int \beta \left( \mathbf{1}_{x \neq y} \frac{dp}{d\mu_0}(y) \right) \mu_0(dy)$  and  $c(x, p) := \int \alpha \left( \int \mathbf{1}_{x \neq y} p(dy) \right)$ ), respectively, as well as the “bar” transport cost with  $c(x, p) := \theta \left( x - \int y p(dy) \right)$ . Our main result is thus to prove a Kantorovich duality theorem for such general transport costs: under some technical assumptions on  $c$  – see below for a precise statement – it holds

$$\mathcal{T}_c(\nu|\mu) = \sup_{\varphi} \left\{ \int R_c \varphi d\mu - \int \varphi d\nu \right\},$$

with

$$R_c\varphi(x) := \inf_p \left\{ \int \varphi dp + c(x, p) \right\}, \quad x \in X,$$

where the supremum (resp. the infimum) above runs over an appropriate class of continuous functions  $\varphi$  (resp. probability measures  $p$ ).

The proofs of such a result and of its corollaries (involving the specific costs mentioned above) are quite involved and rely on a careful choice of the topology on the set of probability measures (with an identification of its dual) together with the Fenchel-Legendre duality theorem, and, for the corollaries, on a careful verification of some technical assumptions.

Besides the applications mentioned above, taking advantage of the above duality result, it is then possible to characterize, extending Bobkov-Götze [6] and the first named author [9], the associated transport-entropy inequality of Talagrand-type (by means of the inf-convolution operator  $R_c$  and using dimension-free concentration properties). Also, specializing to the “bar” cost and extending ideas from [12, 13], the duality result will allow us to characterize the log-Sobolev inequality restricted to convex/concave functions. Examples are provided at the end of the paper.

## CONTENTS

1. Introduction	1
2. Notations	4
2.1. Space and topology	4
2.2. Costs and couplings	5
3. Weak transport costs and their dual forms	5
4. A proof of a result by Strassen	9
5. Characterization of weak transport-entropy inequalities	10
5.1. Bobkov-Götze dual characterization	11
5.2. Tensorisation	12
5.3. Dimension-free concentration	15
6. Weak transport-entropy and log-Sobolev type inequalities	20
6.1. Transport-entropy and $(\tau)$ -log-Sobolev inequalities	21
6.2. Weak transport-entropy inequalities $\overline{\mathcal{T}}_2^\pm$	22
7. Examples	28
7.1. Weak transport inequality for the Bernoulli measure and the product of Bernoulli measures	29
7.2. Weak transport cost for the binomial law	30
7.3. Weak transport cost inequality for the Poisson measure	32
7.4. Sufficient condition for $\overline{\mathbf{T}}_2^-$ on the line	35
7.5. Weak transport cost inequalities with the Hamming distance	35
8. Proof of the duality Theorem and of its corollaries	36
8.1. Fenchel-Legendre duality	36
8.2. Proof of Theorem 3.5 (Duality)	37

8.3. Proofs of Corollaries 3.7, 3.9, 3.11 and 3.12.	41
8.4. Proof of Proposition 3.3	44
Appendix A.	46
References	48

## 2. NOTATIONS

In this section, we introduce the various notations that we use in the rest of the paper.

**2.1. Space and topology.** Throughout the paper  $(X, d)$  is a complete separable metric space. The space of all Borel probability measures on  $X$  is denoted by  $\mathcal{P}(X)$  and the space of all Borel signed measures by  $\mathcal{M}(\mathcal{X})$ . Let  $\gamma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be a lower-semicontinuous function satisfying

$$(2.1) \quad \gamma(0) = 0 \quad \text{and} \quad \gamma(u + v) \leq C(\gamma(u) + \gamma(v)), \quad u, v \in \mathbb{R}_+$$

for some constant  $C$ . We set

$$\mathcal{M}_\gamma(X) := \left\{ \mu \in \mathcal{M}(X); \int \gamma(d(x, x_o)) |\mu|(dx) < \infty \right\}$$

for some (hence all)  $x_o \in X$ . In the applications we shall mainly consider the specific cases  $\gamma_r(u) := u^r$ ,  $u \geq 0$ ,  $r \geq 1$ , and  $\gamma_0(u) := \mathbf{1}_{u \neq 0}$ ,  $u \geq 0$  for which we use the simpler notation  $\mathcal{M}_r(X) := \mathcal{M}_{\gamma_r}(X)$ ,  $r = 0, r \geq 1$ . We equip  $\mathcal{M}_\gamma(X)$  with the coarsest topology that makes continuous the linear functionals  $\mu \mapsto \int \varphi d\mu$ ,  $\varphi \in \Phi_\gamma(X)$ , where  $\Phi_\gamma(X)$  denotes the set of continuous functions  $\varphi : X \rightarrow \mathbb{R}$  satisfying the growth condition

$$(2.2) \quad |\varphi(x)| \leq a + b\gamma(d(x, x_o)), \quad \forall x \in X,$$

for some  $a, b \geq 0$  and some (hence all)  $x_o \in X$ . This topology is denoted by  $\sigma(\mathcal{M}_\gamma(X))$ .

To be more specific, a basis for this topology is given by all finite intersections of sets of the form

$$(2.3) \quad U_{\varphi, a, \varepsilon} := \left\{ m \in \mathcal{M}_\gamma(X); \left| \int \varphi dm - a \right| < \varepsilon \right\}, \quad \varphi \in \Phi_\gamma(X), \quad a \in \mathbb{R}, \quad \varepsilon > 0.$$

The set  $\mathcal{P}_\gamma(X) := \mathcal{P}(X) \cap \mathcal{M}_\gamma(X)$  is equipped with the trace topology denoted by  $\sigma(\mathcal{P}_\gamma(X))$ . For  $\gamma = \gamma_0$ ,  $\Phi_0(X) := \Phi_{\gamma_0}(X) = \mathcal{C}_b(X)$  is the set of all bounded continuous functions and  $\mathcal{P}_0(X) := \mathcal{P}_{\gamma_0}(X) = \mathcal{P}(X)$ . In this case, the topology  $\sigma(\mathcal{P}_0(X))$  is the usual weak topology on  $\mathcal{P}(X)$ .

We define similarly the spaces  $\mathcal{P}_\gamma(X \times X) \subset \mathcal{M}_\gamma(X \times X)$  and equip them with the topologies  $\sigma(\mathcal{M}_\gamma(X \times X))$  and  $\sigma(\mathcal{P}_\gamma(X \times X))$  defined with the class  $\Phi_\gamma(X \times X)$  of continuous functions  $\varphi : X \times X \rightarrow \mathbb{R}$  such that there exist  $a, b \geq 0$  and  $x_o \in X$  such that  $|\varphi(x, y)| \leq a + b(\gamma(d(x_o, x)) + \gamma(d(x_o, y)))$  for all  $x, y \in X$ .

All the spaces considered above will always be equipped with their Borel sigma fields.

Finally, we denote by  $\Phi_{\gamma, b}(X)$ , the set of the elements of  $\Phi_\gamma(X)$  that are bounded from below.

**2.2. Costs and couplings.** A cost is a measurable function  $c: X \times \mathcal{P}_\gamma(X) \rightarrow [0, \infty]$ , for some  $\gamma$ . For all  $\pi \in \mathcal{P}_\gamma(X \times X)$ , we set

$$I_c[\pi] = \int c(x, p_x) \pi_1(dx),$$

where  $\pi_1$  is the first marginal of  $\pi$  and  $x \mapsto p_x$  the ( $\pi_1$  almost everywhere uniquely determined) probability kernel<sup>1</sup> such that

$$\pi(dxdy) = \pi_1(dx)p_x(dy).$$

Note that if  $\pi \in \mathcal{P}_\gamma(X \times X)$ , then  $p_x \in \mathcal{P}_\gamma(X)$  for  $\pi_1$  almost all  $x \in X$  and thus the preceding definition makes sense.

Given two probability measures  $\mu$  and  $\nu$  on  $X$ , we denote by

$$\Pi(\mu, \nu) = \{\pi \in \mathcal{P}(X \times X); \pi(dx \times X) = \mu(dx) \text{ and } \pi(X \times dy) = \nu(dy)\}$$

the set of all couplings  $\pi$  whose first marginal is  $\mu$  and whose second marginal is  $\nu$ . Note also that if,  $\mu, \nu \in \mathcal{P}_\gamma(X)$ , then  $\Pi(\mu, \nu) \subset \mathcal{P}_\gamma(X \times X)$ .

### 3. WEAK TRANSPORT COSTS AND THEIR DUAL FORMS

In this section, we introduce the notion of weak transport costs we shall deal with, some technical assumptions and state a generalization of Kantorovich duality theorem (our first main theorem).

Using the above notations, we introduce a variant of the well-known Monge-Kantorovich optimal transport costs as follows.

**Definition 3.1.** Let  $c: X \times \mathcal{P}_\gamma(X) \rightarrow [0, \infty]$  and  $\mu, \nu \in \mathcal{P}_\gamma(X)$ . The transport cost  $\mathcal{T}_c(\nu|\mu)$  between  $\mu$  and  $\nu$  is defined by

$$\mathcal{T}_c(\nu|\mu) := \inf_{\pi \in \Pi(\mu, \nu)} I_c[\pi] = \inf_{\pi \in \Pi(\mu, \nu)} \int c(x, p_x) \mu(dx).$$

This definition encompasses the following particular cases:

(1) When

$$c(x, p) = \int \omega(x, y) p(dy), \quad x \in X, \quad p \in \mathcal{P}(X),$$

with  $\omega: X \times X \rightarrow [0, \infty]$  a lower-semicontinuous cost function, one recovers the usual Monge-Kantorovich optimal transport cost  $\mathcal{T}_\omega$  defined by

$$\mathcal{T}_\omega(\nu, \mu) = \inf_{\pi \in \Pi(\mu, \nu)} \iint \omega(x, y) \pi(dxdy).$$

(2) The choice

$$c(x, p) = \alpha \left( \int \gamma(d(x, y)) p(dy) \right), \quad x \in X, \quad p \in \mathcal{P}_\gamma(X),$$

with  $\alpha: \mathbb{R}_+ \rightarrow [0, \infty]$  a lower-semicontinuous convex function, yields to another family of transport costs

$$\tilde{\mathcal{T}}_\alpha(\nu|\mu) = \int \alpha \left( \int \gamma(d(x, y)) p(dy) \right) \mu(dx)$$

introduced by Marton in [22] (for  $\gamma = \gamma_0$ , defined below, and therefore  $c(x, p) = \alpha(\int \mathbf{1}_{x \neq y} p(dy))$ ). Note that, in general,  $\tilde{\mathcal{T}}_\alpha$  is not symmetric in  $\mu, \nu$ . If  $\omega(x, y) =$

<sup>1</sup>We recall that a probability kernel is a measurable mapping  $X \ni x \mapsto p_x \in \mathcal{P}(X)$ .

$\alpha(\gamma(d(x, y)))$ , then by Jensen's inequality,  $\tilde{\mathcal{T}}_\alpha(\nu|\mu) \leq \mathcal{T}_\omega(\nu, \mu)$ . Finally, using probabilistic notations, one has

$$\tilde{\mathcal{T}}_\alpha(\nu|\mu) = \inf_{(X, Y)} \mathbb{E}[\alpha(\mathbb{E}[\gamma(d(X, Y))|Y])],$$

where the infimum runs over the set of all couples of random variables  $(X, Y)$  where  $X$  has law  $\mu$  and  $Y$  has law  $\nu$ .

- (3) When  $X \subset \mathbb{R}^m$  (equipped with an arbitrary norm  $\|\cdot\|$ ) is a closed set, another variant is obtained by choosing

$$c(x, p) = \theta\left(x - \int y p(dy)\right), \quad x \in X, \quad p \in \mathcal{P}_1(X),$$

where  $\theta : \mathbb{R}^m \rightarrow [0, \infty]$  is a lower-semicontinuous convex function. The corresponding transport cost is denoted by  $\overline{\mathcal{T}}_\theta$  and defined by

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

We use the notation  $\overline{\mathcal{T}}$  with a bar in reference to the *barycenter* entering its definition. As we shall see below, this last family of transport costs has strong connections with convex functions, and convex ordering of probability measures. In particular, the transport cost corresponding to  $\theta(x) = |x|$ ,  $x \in \mathbb{R}$ , will be involved in a new proof of a result by Strassen on the existence of a martingale with given marginals (see Section 4).

- (4) Let  $\beta : \mathbb{R}_+ \rightarrow [0, +\infty]$  be a lower-semicontinuous convex function and  $\mu_0$  be a reference probability measure on  $X$ . The choice

$$c(x, p) = \int \beta\left(\gamma(d(x, y)) \frac{dp}{d\mu_0}(y)\right) \mu_0(dy), \quad x \in X,$$

if  $p \in \mathcal{P}$  is absolutely continuous with respect to  $\mu_0$  on  $X \setminus \{x\}$ , and  $c(x, p) = +\infty$  otherwise, yields the family of weak transport  $\widehat{\mathcal{T}}_\beta$  defined by

$$\widehat{\mathcal{T}}_\beta(\nu|\mu) = \inf_{\pi \in \Pi(\mu, \nu)} \iint \beta\left(\gamma(d(x, y)) \frac{dp_x}{d\mu_0}(dy)\right) \mu_0(dy) \mu(dx),$$

for all measure  $\mu, \nu \in \mathcal{P}_1(X)$ , absolutely continuous with respect to  $\mu_0$ . The above cost was introduced by Talagrand [34] with  $\gamma_0(d(x, y)) = \mathbf{1}_{x \neq y}$ , to measure the distance from a point  $x \in X$  to a subset  $A \subset X$ , namely he introduced the following pseudo-distance

$$D(x, A) = \inf_{p \in \mathcal{P}(A), p \ll \mu_0 \text{ on } X \setminus \{x\}} \int \beta\left(\mathbf{1}_{x \neq y} \frac{dp}{d\mu_0}(y)\right) \mu_0(dy) = \inf_{p \in \mathcal{P}(A)} c(x, p).$$

(When  $\beta$  is positive except at point 0,  $A = \{x \in X, D(x, A) = 0\}$ .) Talagrand used such a pseudo-distance to obtain concentration results on product spaces ([34, Theorem 4.2.]). It also provides Bernstein type deviation inequalities for supremum of empirical processes as developed in [31, Theorem 1.1.] (to get transport-type inequalities and optimal deviation bounds for supremum of empirical processes).

When  $\alpha(x) = x^p$ ,  $x \geq 0$ , we will use the notation  $\mathcal{T}_p$  and  $\tilde{\mathcal{T}}_p$  to denote the costs above. Accordingly, if  $X = \mathbb{R}^m$  is equipped with a norm  $\|\cdot\|$  and  $\theta(x) = \|x\|^p$ , we will denote the third transport cost by  $\overline{\mathcal{T}}_p$ .

Before stating our main result, we need to introduce some technical assumptions and comment on them.

**Definition 3.2** (Conditions  $(C)$ ,  $(C')$ ,  $(C'')$ ). *Given  $(X, d)$  a complete separable metric space and  $c: X \times \mathcal{P}_\gamma(X) \rightarrow [0, \infty]$  a cost function associated to some lower-semicontinuous function  $\gamma: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  satisfying (2.1), we say the condition  $(C)$  holds if*

(C<sub>1</sub>) *For all  $\mu \in \mathcal{P}_\gamma(X)$ , the function  $\pi \mapsto I_c[\pi]$  is lower-semicontinuous on the set*

$$\Pi(\mu, \cdot) := \{\pi \in \mathcal{P}_\gamma(X \times X); \pi(dx \times X) = \mu(dx)\}.$$

*In other words, for all  $s \geq 0$ , the set  $\{\pi \in \Pi(\mu, \cdot); I_c[\pi] \leq s\}$  is closed for the topology  $\sigma(\mathcal{P}_\gamma(X \times X))$ .*

(C<sub>2</sub>) *The function  $p \mapsto c(x, p)$  is convex for all  $x \in X$ .*

(C<sub>3</sub>) *The function  $(x, p) \mapsto c(x, p)$  is continuous with respect to the product topology.*

(C<sub>4</sub>) *The cost  $c$  is such that if  $\mu \in \mathcal{P}_\gamma(X)$  and  $(p_x)_{x \in X}$  are measurable probability kernels such that  $p_x \in \mathcal{P}_\gamma(X)$  for all  $x \in X$  and  $\int c(x, p_x) \mu(dx) < \infty$ , then  $\nu = \mu p \in \mathcal{P}_\gamma(X)$ .*

*Similarly we say that condition  $(C')$  holds if  $(C_1), (C_2), (C_4)$  hold together with*

(C'<sub>3</sub>)  *$(X, d)$  is compact and the function  $(x, p) \mapsto c(x, p)$  is lower-semicontinuous with respect to the product topology,*

*and that condition  $(C'')$  holds if  $(C_2), (C_4)$  hold together with*

(C''<sub>3</sub>)  *$X$  is a countable set of isolated points and for all  $x \in X$ , the function  $p \mapsto c(x, p)$  is lower-semicontinuous.*

The above conditions are technical. However, Condition  $(C_2)$  is the least we can hope for.

As for applications, the main difficulty is coming from Condition  $(C_1)$ . Let us make some comments about this assumption. First specializing to  $\mu = \delta_x$ , condition  $(C_1)$  implies that for all  $x \in X$ , the function  $p \mapsto c(x, p)$  is lower semicontinuous on  $\mathcal{P}_\gamma(X)$ . In the discrete setting, the converse is also true: as shown in the proof of Theorem 3.5, Condition  $(C''_3)$  implies Condition  $(C_1)$  (this is why the latter does not appear in Condition  $(C'')$ ). For more general spaces, we do not know if Condition  $(C_1)$  is strictly stronger than lower-semicontinuity of the cost function  $c$ . Nevertheless, we have the following rather general abstract result whose proof is postponed to Section 8. In particular, such a result applies to the transport costs  $\tilde{T}, \bar{T}$  and  $\hat{T}$ .

**Proposition 3.3.** *Let  $(X, d)$  be complete separable metric space. Let  $(\varphi_k)_{k \in \mathbb{N}}$  be a sequence of elements of  $\Phi_\gamma(X \times X)$  (with  $\gamma: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  satisfying (2.1)) such that  $\varphi_0 \equiv 0$ . Assume that the cost function  $c: X \times \mathcal{P}_\gamma(X) \rightarrow [0, \infty]$  is defined by*

$$(3.4) \quad c(x, p) = \sup_{k \in \mathbb{N}} \int \varphi_k(x, y) p(dy), \quad \forall x \in X, \quad \forall p \in \mathcal{P}_\gamma(X).$$

*Then Conditions  $(C_1)$  and  $(C_2)$  hold and  $c: X \times \mathcal{P}_\gamma(X) \rightarrow [0, \infty]$  is lower-semicontinuous with respect to the product topology.*

We are now in a position to state the main result of this section: a generalization of the Kantorovich duality theorem.

**Theorem 3.5.** *Let  $(X, d)$  be a complete separable metric space. Let  $c: X \times \mathcal{P}_\gamma(X) \rightarrow [0, \infty]$  be a cost function associated to some lower-semicontinuous function  $\gamma: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  satisfying (2.1). Assume that condition  $(C), (C')$  or  $(C'')$  holds. Then, for all  $\mu, \nu \in \mathcal{P}_\gamma(X)$ , the following duality formula holds:*

$$\mathcal{T}_c(\nu|\mu) = \sup_{\varphi \in \Phi_{\gamma, b}(X)} \left\{ \int R_c \varphi(x) \mu(dx) - \int \varphi(y) \nu(dy) \right\},$$



where

$$R_c\varphi(x) := \inf_{p \in \mathcal{P}_\gamma(X)} \left\{ \int \varphi(y) p(dy) + c(x, p) \right\}, \quad x \in X, \quad \varphi \in \Phi_{\gamma, b}(X).$$

**Remark 3.6.** Note that since  $c \geq 0$ ,  $R_c\varphi$  is bounded from below as soon as  $\varphi$  is bounded from below. Therefore,  $\int R_c\varphi(x) \mu(dx)$  is always well defined in  $(-\infty, \infty]$ . Note also, that  $R_c\varphi$  is always measurable. This is clear under Condition  $(C_3)$ , since in this case  $R_c\varphi$  is lower-semicontinuous as an infimum of continuous functions. Under Condition  $(C'_3)$ , it is not difficult to check that  $R_c\varphi$  remains lower-semicontinuous, using the fact that  $\mathcal{P}_\gamma(X)$  is compact.

The proof of Theorem 3.5 together with the forthcoming corollaries are postponed to Section 8 for clarity of exposition.

Specializing Theorem 3.5 to the cost functions introduced above, we obtain the following corollaries. The first result is in fact the usual Kantorovich duality theorem that we recall for the sake of completeness (see for instance [37, Theorem 5.10] and the references therein).

**Corollary 3.7.** Let  $(X, d)$  be a complete separable metric space. Assume that  $\omega : X \times X \rightarrow [0, \infty]$  is some lower-semicontinuous cost function. Then it holds,

$$(3.8) \quad \mathcal{T}_\omega(\nu, \mu) = \sup_{\varphi \in \mathcal{C}_b(X)} \left\{ \int Q_\omega\varphi(x) \mu_*(dx) - \int \varphi(y) \nu(dy) \right\}, \quad \mu, \nu \in \mathcal{P}(X),$$

where  $\mu_*$  denotes the inner measure induced by  $\mu$  and  $Q_\omega\varphi(x) = \inf_{y \in X} \{\varphi(y) + \omega(x, y)\}$ ,  $x \in X$ ,  $\varphi \in \mathcal{C}_b(X)$ .

For transport costs of the form  $\tilde{\mathcal{T}}$ , one has the following result.

**Corollary 3.9.** Assume either that  $(X, d)$  is a complete separable metric space,  $\alpha : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is a convex continuous function with  $\alpha(0) = 0$  and  $\gamma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is continuous, or  $(X, d)$  is either a compact space or a countable set of isolated points,  $\alpha : \mathbb{R}_+ \rightarrow [0, +\infty]$  is a convex lower-semicontinuous function with  $\alpha(0) = 0$  and  $\gamma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is lower-semicontinuous. Then, the following duality formula holds

$$(3.10) \quad \tilde{\mathcal{T}}_\alpha(\nu|\mu) = \sup_{\varphi \in \Phi_{\gamma, b}(X)} \left\{ \int \tilde{Q}_\alpha\varphi(x) \mu(dx) - \int \varphi(y) \nu(dy) \right\}, \quad \mu, \nu \in \mathcal{P}_\gamma(X),$$

where

$$\tilde{Q}_\alpha\varphi(x) = \inf_{p \in \mathcal{P}_\gamma(X)} \left\{ \int \varphi(y) p(dy) + \alpha \left( \int \gamma(d(x, y)) p(dy) \right) \right\}, \quad x \in X, \quad \varphi \in \Phi_{\gamma, b}(X).$$

We observe that, anticipating the present paper, the duality formula (3.10) was already put to use in [15], in connection with displacement convexity of the relative entropy functional on graphs.

As for the ‘‘bar’’ transport cost, the duality formula for  $\overline{\mathcal{T}}_\theta$  can be expressed using convex functions only. This fact will repeatedly be used in the applications.

**Corollary 3.11.** Let  $X \subset \mathbb{R}^m$  be a closed subset of  $\mathbb{R}^m$  equipped with a norm  $\|\cdot\|$  and  $\theta : \mathbb{R}^m \rightarrow \mathbb{R}_+$  be a convex function such that  $\theta(x) \geq a\|x\| + b$ , for all  $x \in \mathbb{R}^m$  and for some  $a > 0$  and  $b \in \mathbb{R}$ . Then,

(1) The following duality identity holds

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \sup_{\varphi \in \Phi_{1, b}(X)} \left\{ \int \overline{Q}_\theta\varphi(x) \mu(dx) - \int \varphi(y) \nu(dy) \right\}, \quad \mu, \nu \in \mathcal{P}_1(X),$$

where for all  $x \in \mathbb{R}^m$  and all  $\varphi \in \Phi_{1,b}(X)$ ,

$$\overline{Q}_\theta \varphi(x) = \inf_{p \in \mathcal{P}_1(X)} \left\{ \int \varphi(y) p(dy) + \theta \left( x - \int y p(dy) \right) \right\}.$$

Since  $\mathcal{P}_1(X) \subset \mathcal{P}_1(\mathbb{R}^m)$ , the same conclusion holds replacing  $\Phi_{1,b}(X)$  by  $\Phi_{1,b}(\mathbb{R}^m)$  in the dual expression of  $\overline{T}_\theta(\nu|\mu)$  and  $\mathcal{P}_1(X)$  by  $\mathcal{P}_1(\mathbb{R}^m)$  in the definition of  $\overline{Q}_\theta \varphi$ .  
 (2) For all  $\varphi \in \Phi_{1,b}(\mathbb{R}^m)$  and all  $x \in \mathbb{R}^m$ , it holds

$$\overline{Q}_\theta \varphi(x) := \inf_{p \in \mathcal{P}_1(\mathbb{R}^m)} \left\{ \int \varphi(y) p(dy) + \theta \left( x - \int y p(dy) \right) \right\} = Q_\theta \overline{\varphi}(x),$$

where  $\overline{\varphi}$  denotes the greatest convex function  $h : \mathbb{R}^m \rightarrow \mathbb{R}$  such that  $h \leq \varphi$ , and we recall that  $Q_\theta g(x) = \inf_{y \in \mathbb{R}^m} \{g(y) + \theta(x - y)\}$ ,  $g \in \Phi_{1,b}(\mathbb{R}^m)$ ,  $x \in \mathbb{R}^m$ .

(3) For all  $\mu, \nu \in \mathcal{P}_1(X)$ , it holds

$$\overline{T}_\theta(\nu|\mu) = \sup \left\{ \int Q_\theta \varphi d\mu - \int \varphi d\nu; \varphi : \mathbb{R}^m \rightarrow \mathbb{R}, \text{ convex, Lipschitz, bounded from below} \right\}$$

(4) In particular, if  $\theta(x) = \|x\|$ ,  $x \in \mathbb{R}^m$ , then for all  $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^m)$ ,

$$\overline{T}_1(\nu|\mu) = \sup \left\{ \int \varphi d\mu - \int \varphi d\nu; \varphi \text{ convex, 1-Lipschitz, bounded from below} \right\}.$$

The results (1), (2), (3) also hold when  $\theta : \mathbb{R}^m \rightarrow [0, +\infty]$  is a lower semi-continuous convex function and  $X$  is either compact or a countable set of isolated points.

Finally we state a duality theorem for the ‘‘hat’’ transport cost.

**Corollary 3.12.** *Let  $(X, d)$  be a compact metric space or a countable set of isolated points. Let  $\beta : \mathbb{R}_+ \rightarrow [0, +\infty]$  be a lower-semicontinuous convex function with  $\beta(0) = 0$  and  $\lim_{x \rightarrow \infty} \beta(x)/x = +\infty$ . Assume that  $\gamma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is lower-semicontinuous with  $\gamma(0) = 0$  and  $\gamma(u) > 0$  for all  $u > 0$ . Let  $\mu_0$  be a reference probability measure on  $X$ . Then, for all  $\mu, \nu \in \mathcal{P}_\gamma(X)$  absolutely continuous with respect to  $\mu_0$ , it holds*

$$\widehat{T}_\beta(\nu|\mu) = \sup_{\varphi \in \Phi_{\gamma,b}(X)} \left\{ \int \widehat{Q}_\beta \varphi(x) \mu(dx) - \int \varphi(y) \nu(dy) \right\},$$

where for  $x \in X$  and  $\varphi \in \Phi_{\gamma,b}(X)$ ,

$$\widehat{Q}_\beta \varphi(x) := \inf_{p \in \mathcal{P}_\gamma(X), p \ll \mu_0 \text{ on } X \setminus \{x\}} \left\{ \int \varphi(y) p(dy) + \int \beta \left( \gamma(d(x, y)) \frac{dp}{d\mu_0}(y) \right) d\mu_0(y) \right\}.$$

#### 4. A PROOF OF A RESULT BY STRASSEN

In this short section, we show that the transport cost  $\overline{T}_\theta$  can be used to recover an old result by Strassen [33] about the existence of a martingale with given marginals.

In the sequel, we equip  $\mathbb{R}^m$  with an arbitrary norm  $\|\cdot\|$ . Let  $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^m)$ ; one says that  $\mu$  is dominated by  $\nu$  in the convex order sense, and one writes  $\mu \preceq_C \nu$ , if

$$\int f d\mu \leq \int f d\nu,$$

for all convex<sup>2</sup>  $f : \mathbb{R}^m \rightarrow \mathbb{R}$ . Note that, in particular, this implies that  $\int f d\mu = \int f d\nu$  for all affine maps  $f : \mathbb{R}^m \rightarrow \mathbb{R}$ .

<sup>2</sup>Note that since  $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^m)$ , any affine map is integrable with respect to  $\mu$  and  $\nu$ . Since a convex function is always positive up to the addition of some affine map, we see that the integral of convex functions with respect to  $\mu$  and  $\nu$  makes sense.

It is not difficult to check that  $\mu \preceq_C \nu$  if and only if  $\int f d\mu \leq \int f d\nu$  for all 1-Lipschitz and convex  $f: \mathbb{R}^m \rightarrow \mathbb{R}$  bounded from below<sup>3</sup>.

The following result goes back at least to the work of Strassen [33].

**Theorem 4.1.** *Let  $\mu, \nu \in \mathcal{P}(\mathbb{R}^m)$ ; there exists a martingale  $(X, Y)$ , where  $X$  follows the law  $\mu$  and  $Y$  the law  $\nu$  if and only if  $\mu \preceq_C \nu$ .*

*Proof.* If  $\pi \in \Pi(\mu, \nu)$  denotes the law of  $(X, Y)$ , the condition that  $(X, Y)$  is a martingale is expressed by

$$(4.2) \quad \int y p_x(dy) = x, \quad \text{for } \mu \text{ almost every } x \in \mathbb{R}^m.$$

Recall that  $\overline{\mathcal{T}}_1(\nu|\mu) = \inf_{\pi \in \Pi(\mu, \nu)} \int \|x - \int y p_x(dy)\| \mu(dx)$ . Therefore, there exists some  $\pi \in \Pi(\mu, \nu)$  satisfying (4.2) if and only if  $\overline{\mathcal{T}}_1(\nu|\mu) = 0$ . Since, by Corollary 3.11,  $\overline{\mathcal{T}}_1(\nu|\mu) = \sup \{ \int f d\mu - \int f d\nu; f: \mathbb{R}^m \rightarrow \mathbb{R}, 1\text{-Lipschitz, convex and bounded below} \}$ , the expected result follows.  $\square$

## 5. CHARACTERIZATION OF WEAK TRANSPORT-ENTROPY INEQUALITIES

In this section, we introduce a general notion of (weak) transport-entropy inequalities of Talagrand-type and investigate them. We recall that if  $\mu, \nu$  are two probability measures on some space  $X$ , the relative entropy of  $\nu$  with respect to  $\mu$  is defined by

$$H(\nu|\mu) = \int \log \left( \frac{d\nu}{d\mu} \right) d\nu \in \mathbb{R}_+ \cup \{+\infty\},$$

if  $\nu \ll \mu$ . Otherwise, one sets  $H(\nu|\mu) = +\infty$ .

**Definition 5.1** (Transport-entropy inequalities  $\mathbf{T}_c(a_1, a_2)$  and  $\mathbf{T}_c(b)$ ). *Let  $c: X \times \mathcal{P}_\gamma(X) \rightarrow [0, \infty]$  be a cost function associated to some lower-semicontinuous function  $\gamma: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  satisfying (2.1). The probability measure  $\mu \in \mathcal{P}_\gamma(X)$  is said to satisfy  $\mathbf{T}_c(a_1, a_2)$ , for some  $a_1, a_2 > 0$  if for all  $\nu_1, \nu_2 \in \mathcal{P}_\gamma(X)$  it holds*

$$(5.2) \quad \mathcal{T}_c(\nu_1|\nu_2) \leq a_1 H(\nu_1|\mu) + a_2 H(\nu_2|\mu).$$

*On the other hand,  $\mu \in \mathcal{P}_\gamma(X)$  is said to satisfy  $\mathbf{T}_c^+(b)$ , respectively  $\mathbf{T}_c^-(b)$ , for some  $b > 0$ , if for all  $\nu \in \mathcal{P}_\gamma(X)$  it holds*

$$(5.3) \quad \mathcal{T}_c(\nu|\mu) \leq b H(\nu|\mu),$$

respectively

$$(5.4) \quad \mathcal{T}_c(\mu|\nu) \leq b H(\nu|\mu).$$

*For the specific transport costs  $\tilde{\mathcal{T}}_p$  and  $\overline{\mathcal{T}}_p$  introduced in Section 3 (with  $\alpha(x) = x^p, x \geq 0$  and  $\theta(x) = \|x\|^p, x \in \mathbb{R}^m, p \geq 1$ ), we may use the corresponding notations  $\tilde{\mathbf{T}}_p(a_1, a_2)$ ,  $\tilde{\mathbf{T}}_p^\pm(b)$ , respectively  $\overline{\mathbf{T}}_p(a_1, a_2)$ ,  $\overline{\mathbf{T}}_p^\pm(b)$ .*

Let us comment on this definition. First we note, that when  $c(x, p) = \int \omega(x, y) p(dy)$ , (5.3) and (5.4) give back the usual transport-entropy inequalities of Talagrand type (see [18], [37] or [10] for a general introduction on the subject). Also, we observe that  $\mathbf{T}_c(a_1, 0)$  or  $\mathbf{T}_c(a_2, 0)$  (which are not considered in the above definition, since  $a_1, a_2 > 0$ ) has no meaning. Indeed, if  $\mathbf{T}_c(a_1, 0)$  holds, then  $\mathcal{T}_c(\nu_1|\nu_2) \leq a_1 H(\nu_1|\mu)$  for all  $\nu_1, \nu_2$  which in turn implies  $\mathcal{T}_c(\mu|\nu_2) = 0$  for all  $\nu_2$  which is impossible. Finally, using the convention

<sup>3</sup>One possible way to prove this is to use the fact that if  $f: \mathbb{R}^m \rightarrow \mathbb{R}$  is convex, then the classical inf-convolution operator  $Q_t f(x) := \inf_{y \in \mathbb{R}^m} \{f(y) + \frac{1}{t} \|x - y\|\}$  is convex,  $1/t$ -Lipschitz and  $Q_t f(x) \uparrow f(x)$  when  $t \rightarrow 0$  for all  $x \in \mathbb{R}^m$ .

that  $0 \cdot \infty = 0$ , we observe that  $\mathbf{T}_c^+(b)$  is formally equivalent to  $\mathbf{T}_c(b, \infty)$ , and  $\mathbf{T}_c^-(b)$  is equivalent to  $\mathbf{T}_c^-(\infty, b)$ .

As for the classical inequality,  $\mathbf{T}_c(a_1, a_2)$  does enjoy the tensorisation property. Moreover, using the dual formulation of Section 3, we can state two different characterizations of  $\mathbf{T}_c(a_1, a_2)$ : one in the spirit of Bobkov-Götze dual formulation, and one in the spirit of the first named author's characterization of dimension-free concentration property [9]. We now state these properties and characterizations.

**5.1. Bobkov-Götze dual characterization.** The following characterization extends, thanks to the dual formulation of the transport cost [6]; see also [10].

**Proposition 5.5** (Dual formulation). *Let  $c: X \times \mathcal{P}_\gamma(X) \rightarrow [0, \infty]$  be a cost function associated to some lower-semicontinuous function  $\gamma: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  satisfying (2.1). Assume that  $c(x, \delta_x) = 0$  for all  $x \in X$  and that Condition (C), (C') or (C'') holds. For  $\mu \in \mathcal{P}_\gamma(X)$  and  $a_1, a_2, b > 0$ , Items (i)'s and (ii)'s are equivalent:*

$$(5.6) \quad \begin{aligned} & \bullet \text{ (i) } \mathbf{T}_c(a_1, a_2) \text{ holds;} \\ & \bullet \text{ (ii) for all } \varphi \in \Phi_{\gamma, b}(X) \text{ (resp. for all non-negative } \varphi \in \Phi_\gamma), \text{ it holds} \\ & \left( \int \exp \left\{ \frac{R_c \varphi}{a_2} \right\} d\mu \right)^{a_2} \left( \int \exp \left\{ -\frac{\varphi}{a_1} \right\} d\mu \right)^{a_1} \leq 1; \end{aligned}$$

$$(5.7) \quad \begin{aligned} & \bullet \text{ (i') } \mathbf{T}_c^+(b) \text{ holds;} \\ & \bullet \text{ (ii') for all } \varphi \in \Phi_{\gamma, b}(X) \text{ (resp. for all non-negative } \varphi \in \Phi_\gamma), \text{ it holds} \\ & \exp \left\{ \int R_c \varphi d\mu \right\} \left( \int \exp \left\{ \frac{-\varphi}{b} \right\} d\mu \right)^b \leq 1; \end{aligned}$$

$$(5.8) \quad \begin{aligned} & \bullet \text{ (i'') } \mathbf{T}_c^-(b) \text{ holds;} \\ & \bullet \text{ (ii'') for all } \varphi \in \Phi_{\gamma, b}(X) \text{ (resp. for all non-negative } \varphi \in \Phi_\gamma), \text{ it holds} \\ & \left( \int \exp \left\{ \frac{R_c \varphi}{b} \right\} d\mu \right)^b \exp \left\{ - \int \varphi d\mu \right\} \leq 1, \end{aligned}$$

where we recall that  $R_c \varphi(x) = \inf_{p \in \mathcal{P}_\gamma(X)} \{ \int \varphi(y) p(dy) + c(x, p) \}$ ,  $x \in X$ .

Moreover, specializing to the “bar” cost  $\overline{T}_\theta$ , one can replace, in (ii), (ii') and (ii''),  $R_c \varphi$  by  $Q_\theta \varphi := \inf_{y \in \mathbb{R}^m} \{ \varphi(y) + \theta(\cdot - y) \}$  and restrict to the set of functions  $\varphi$  that are convex, Lipschitz and bounded from below.

**Remark 5.9.** When  $c(x, p) = \theta(x - \int y p(dy))$ ,  $x \in \mathbb{R}^m$ ,  $p \in \mathcal{P}_1(\mathbb{R}^m)$ , for some convex function  $\theta: \mathbb{R}^m \rightarrow \mathbb{R}_+$ , the inequality  $\mathbf{T}_c(a_1, a_2)$  is thus equivalent to the following exponential type inequality first introduced by Maurey [24] (the so-called convex ( $\tau$ )-property):

$$\left( \int e^{\frac{Q_\theta \varphi}{a_2}} d\mu \right)^{a_2} \left( \int e^{-\frac{\varphi}{a_1}} d\mu \right)^{a_1} \leq 1, \quad \forall \varphi: \mathbb{R}^m \rightarrow \mathbb{R}_+ \text{ convex.}$$

*Proof.* By duality (i.e. using Theorem 3.5),  $\mathbf{T}_c(a_1, a_2)$  is equivalent to have

$$a_2 \left( \int \frac{R_c \varphi}{a_2} d\nu_2 - H(\nu_2 | \mu) \right) + a_1 \left( \int -\frac{\varphi}{a_1} d\nu_1 - H(\nu_1 | \mu) \right) \leq 0,$$

for all  $\varphi \in \Phi_{\gamma, b}(X)$  and all  $\nu_1, \nu_2 \in \mathcal{P}_\gamma(X)$  with finite relative entropy with respect to  $\mu$ . The expected result follows by taking the (two independent) suprema, on the left hand side, over  $\nu_1$  and  $\nu_2$ , and by using Lemma 5.10 below. Note that since  $c(x, \delta_x) = 0$  for all  $x \in X$ , one always has  $R_c \varphi \leq \varphi$ , for all  $\varphi \in \Phi_{\gamma, b}(X)$  and so the function  $\psi = R_c \varphi / a_2$  satisfies the assumption of Lemma 5.10. This completes the proof of the equivalence (i)  $\Leftrightarrow$  (ii).

Note that (5.6) is invariant under translations  $\varphi \mapsto \varphi + a$  and so the functions  $\varphi$  can be assumed non-negative.

The two last equivalences follow the same line (and the details are left to the reader). Similarly, the specialization to the “bar” cost is identical, one just needs to apply Item (3) of Corollary 3.11 rather than Theorem 3.5.  $\square$

**Lemma 5.10.** *Let  $\mu \in \mathcal{P}_\gamma(X)$  for some lower-semicontinuous function  $\gamma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  satisfying (2.1); for all measurable function  $\psi : \mathcal{X} \rightarrow \mathbb{R}$  such that  $\psi \leq \varphi$  for some  $\varphi \in \Phi_\gamma(X)$ , it holds*

$$\sup_{\nu \in \mathcal{P}_\gamma(X)} \left\{ \int \psi d\nu - H(\nu|\mu) \right\} = \log \int e^\psi d\mu.$$

*Proof of Lemma 5.10.* Consider the function  $U(x) = x \log(x)$ ,  $x > 0$ . A simple calculation shows that  $U^*(t) := \sup_{x>0} \{tx - U(x)\} = e^{t-1}$ ,  $t \in \mathbb{R}$ . Since  $\psi \leq \varphi$ , for some  $\varphi \in \Phi_\gamma(X)$ , one concludes that  $\int [\psi]_+ d\nu$  is finite for all  $\nu \in \mathcal{P}_\gamma(X)$ , and thus  $\int \psi d\nu$  is well-defined in  $\mathbb{R} \cup \{-\infty\}$ . Let  $\nu \ll \mu$ ; applying Young’s inequality  $xy \leq U(x) + U^*(y)$ ,  $x > 0, y \in \mathbb{R}$ , one gets

$$\int \psi d\nu \leq \int U^*(\psi) d\mu + \int U\left(\frac{d\nu}{d\mu}\right) d\mu = \int e^{\psi-1} d\mu + H(\nu|\mu).$$

Applying this inequality to  $\psi + u$ , where  $u \in \mathbb{R}$ , we get

$$\int \psi d\nu - H(\nu|\mu) \leq e^{u-1} \int e^\psi d\mu - u,$$

and this inequality is still true, even if  $\nu$  is not absolutely continuous with respect to  $\mu$ . Optimizing over  $u \in \mathbb{R}$  and over  $\nu \in \mathcal{P}_\gamma(X)$  yields:

$$\sup_{\nu \in \mathcal{P}_\gamma(X)} \left\{ \int \psi d\nu - H(\nu|\mu) \right\} \leq \log \int e^\psi d\mu.$$

To get the converse inequality, consider, for  $k \geq 0$  large enough,  $\nu_k(dx) = \frac{e^{\psi(x)}}{\int e^\psi \mathbf{1}_{A_k} d\mu} \mathbf{1}_{A_k}(x) \mu(dx)$ , where  $A_k = \{x \in X; \psi(x) \leq k\}$ . Since  $\mu$  belongs to  $\mathcal{P}_\gamma(X)$  and  $\nu_k$  has a bounded density with respect to  $\mu$ ,  $\nu_k$  also belongs to  $\mathcal{P}_\gamma(X)$ . Furthermore

$$\int \psi d\nu_k - H(\nu_k|\mu) = \log \left( \int e^\psi \mathbf{1}_{A_k} d\mu \right) \rightarrow \log \left( \int e^\psi d\mu \right),$$

when  $k \rightarrow \infty$ . This completes the proof.  $\square$

**5.2. Tensorisation.** In this section, we collect two important properties which will allow us to deal with one-dimensional measures in applications.

**Theorem 5.11** (Tensorisation property). *Let  $\gamma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be a lower-semicontinuous function satisfying (2.1),  $(X_1, d_1), \dots, (X_n, d_n)$  be complete separable metric spaces equipped with cost functions  $c_i : X_i \times \mathcal{P}_\gamma(X_i) \rightarrow [0, \infty]$ ,  $i \in \{1, \dots, n\}$  such that  $c_i(x_i, \delta_{x_i}) = 0$  and  $p_i \mapsto c_i(x_i, p_i)$  is convex for all  $x_i \in X_i$ . For all  $i \in \{1, \dots, n\}$ , let  $\mu_i \in \mathcal{P}_\gamma(X_i)$  satisfying the transport inequality  $\mathbf{T}_{c_i}(a_1^{(i)}, a_2^{(i)})$  for some  $a_1^{(i)}, a_2^{(i)} > 0$ . Then the product probability measure  $\mu_1 \otimes \dots \otimes \mu_n$  satisfies the transport inequality  $\mathbf{T}_c(a_1, a_2)$ , with  $a_1 := \max_i a_1^{(i)}$ ,  $a_2 := \max_i a_2^{(i)}$ , for the cost function  $c : X_1 \times \dots \times X_n \times \mathcal{P}_\gamma(X_1 \times \dots \times X_n) \rightarrow [0, \infty)$  defined by*

$$c(x, p) = c_1(x_1, p_1) + \dots + c_n(x_n, p_n),$$

for all  $x = (x_1, \dots, x_n) \in X_1 \times \dots \times X_n$ , and for all  $p \in \mathcal{P}_\gamma(X_1 \times \dots \times X_n)$ , where  $p_i$  denotes the  $i$ -th marginal distribution of  $p$ .

The following is an immediate corollary of Theorem 5.11.

**Corollary 5.12.** *Let  $\gamma: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be a lower-semicontinuous function satisfying (2.1) and assume that  $\mu \in \mathcal{P}_\gamma(X)$  satisfies the transport inequality  $\mathbf{T}_c(a_1, a_2)$  for some  $a_1, a_2 > 0$  and some cost function  $c: X \times \mathcal{P}_\gamma(X) \rightarrow [0, \infty]$  that satisfies  $c(x, \delta_x) = 0$  and  $p \mapsto c(x, p)$  convex for all  $x \in X$ . Then for all positive integers  $n$ , the product probability measure  $\mu^n \in \mathcal{P}_\gamma(X^n)$  satisfies the inequality  $\mathbf{T}_{c^n}(a_1, a_2)$ , where  $c^n: X^n \times \mathcal{P}_\gamma(X^n) \rightarrow [0, \infty]$  is the cost function defined by*

$$c^n(x, p) := \sum_{i=1}^n c(x_i, p_i), \quad x = (x_1, \dots, x_n) \in X^n, \quad p \in \mathcal{P}_\gamma(X^n),$$

where  $p_i$  denotes the  $i$ -th marginal distribution of  $p$ .

The proof of Theorem 5.11 uses the chain rule formula for the entropy on the one hand, and on the other, a similar property for the transport cost, which we now state in the following lemma of independent interest.

**Lemma 5.13** (Chain rule inequality for the transport cost). *Let  $\gamma: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be a lower-semicontinuous function satisfying (2.1),  $(X_1, d_1), (X_2, d_2)$  be complete separable metric spaces equipped with cost functions  $c_i: X_i \times \mathcal{P}_\gamma(X_i) \rightarrow [0, \infty]$ ,  $i \in \{1, 2\}$  such that  $c_i(x_i, \delta_{x_i}) = 0$  and  $p_i \mapsto c_i(x_i, p_i)$  is convex for all  $x_i \in X_i$ . Define  $c: X_1 \times X_2 \times \mathcal{P}_\gamma(X_1 \times X_2) \rightarrow [0, \infty]$  by  $c(x, p) = c_1(x_1, p_1) + c_2(x_2, p_2)$ ,  $x = (x_1, x_2) \in X_1 \times X_2$ ,  $p \in \mathcal{P}_\gamma(X_1 \times X_2)$ , where  $p_i$  denotes the  $i$ -th marginal distribution of  $p$ .*

Then, for all  $\nu, \nu' \in \mathcal{P}_\gamma(X_1 \times X_2)$ , all  $\varepsilon > 0$ , there exists a kernel  $p_1^\varepsilon$  such that

$$\mathcal{T}_c(\nu'|\nu) \leq \mathcal{T}_{c_1}(\nu'_1|\nu_1) + \int_{X_1 \times X_1} \mathcal{T}_{c_2}(\nu'_2(y_1, \cdot)|\nu_2(x_1, \cdot)) p_1^\varepsilon(x_1, dy_1) \nu_1(dx_1) + 2\varepsilon,$$

where  $\nu_1$  and  $\nu'_1$  are the first marginals of  $\nu, \nu'$  respectively; the kernels  $x_1 \mapsto \nu_2(x_1, \cdot)$  and  $y_1 \mapsto \nu'_2(y_1, \cdot)$  are such that

$$\nu(dx_1 dx_2) = \nu_1(dx_1) \nu_2(x_1, dx_2) \quad \text{and} \quad \nu'(dy_1 dy_2) = \nu'_1(dy_1) \nu'_2(y_1, dy_2);$$

and the kernel  $p_1^\varepsilon$ , defined so that  $\pi_1^\varepsilon(dx_1 dy_1) := \nu_1(dx_1) p_1^\varepsilon(x_1, dy_1) \in \Pi(\nu_1, \nu'_1)$ , satisfies  $\mathcal{T}_{c_1}(\nu_1|\nu'_1) \geq \int_{X_1 \times X_1} c_1(x_1, p_1^\varepsilon(x_1, \cdot)) \nu_1(dx_1) - \varepsilon$ .

**Remark 5.14.** *If one assumes that the cost functions  $c_1$  and  $c_2$  satisfy assumption  $(C_1)$ , then the error term  $\varepsilon$  can be chosen 0. Indeed, under assumption  $(C_1)$  the function  $\pi \mapsto \int c_1(x, p_x) \nu'_1(dx_1)$  is lower semicontinuous on the set  $\Pi(\nu'_1, \nu_1)$  which is easily seen to be compact (using Theorem 8.3 below). Therefore it attains its infimum and so there exists some kernel  $p_1$  such that  $\mathcal{T}_{c_1}(\nu_1|\nu'_1) = \int c_1(x_1, p_1(x_1, \cdot)) \nu'_1(dx_1)$ . The same applies for cost functions based on the cost  $c_2$ .*

*Proof of Lemma 5.13.* Fix  $\nu, \nu' \in \mathcal{P}_\gamma(X_1 \times X_2)$  and  $\varepsilon > 0$ . Our aim is first to define a probability kernel  $p$  appropriately related to  $\nu$  and  $\nu'$ .

To that purpose, let  $p_1$  be a probability kernel (that depends on  $\varepsilon$  although not explicitly stated for simplicity) so that  $\pi_1(dx_1 dy_1) := \nu_1(dx_1) p_1(x_1, dy_1) \in \Pi(\nu_1, \nu'_1)$  and

$$(5.15) \quad \int_{X_1 \times X_1} c_1(x_1, p_1(x_1, \cdot)) \nu_1(dx_1) \leq \mathcal{T}_{c_1}(\nu'_1|\nu_1) + \varepsilon.$$

Similarly, for all  $x_1, y_1 \in X_1$ , let  $X_2 \ni x_2 \mapsto q_2^{x_1, y_1}(x_2, \cdot) \in \mathcal{P}(X_2)$  be a probability kernel (that depends also on  $\varepsilon$ ) satisfying  $\pi_2^{x_1, y_1}(dx_2 dy_2) := \nu_2(x_1, dx_2) q_2^{x_1, y_1}(x_2, dy_2) \in \Pi(\nu_2(x_1, \cdot), \nu'_2(y_1, \cdot))$  and

$$(5.16) \quad \int_{X_2 \times X_2} c_2(x_2, q_2^{x_1, y_1}(x_2, \cdot)) \nu_2(x_1, dx_2) \leq \mathcal{T}_{c_2}(\nu'_2(y_1, \cdot)|\nu_2(x_1, \cdot)) + \varepsilon.$$

Then observe that for all  $f : X_1 \times X_2 \rightarrow \mathbb{R}$  it holds

$$\begin{aligned} & \int f(y_1, y_2) p_1(x_1, dy_1) q_2^{x_1, y_1}(x_2, dy_2) \nu(dx_1 dx_2) \\ &= \int f(y_1, y_2) p_1(x_1, dy_1) q_2^{x_1, y_1}(x_2, dy_2) \nu_2(x_1, dx_2) \nu_1(dx_1) \\ &= \int f(y_1, y_2) p_1(x_1, dy_1) \nu'_2(y_1, dy_2) \nu_1(dx_1) = \int f(y_1, y_2) \nu'_2(y_1, dy_2) \nu'_1(dy_1) = \int f(y) \nu'(dy). \end{aligned}$$

Hence,  $p(x, dy) := p_1(x_1, dy_1) q_2^{x_1, y_1}(x_2, dy_2)$  is a probability kernel satisfying  $\pi(dxdy) := p(x, dy) \nu(dx) \in \Pi(\nu, \nu')$ . Let  $p_2(x, \cdot) := \int_{X_1} p_1(x_1, dy_1) q_2^{x_1, y_1}(x_2, \cdot) \in \mathcal{P}(X_2)$  be the second marginal of  $p(x_1, \cdot)$ , observing that  $p_1(x, \cdot)$  is its first marginal.

Finally, using the definition of the transport cost, the definition of the cost and Jensen's inequality, it holds

$$\begin{aligned} \mathcal{T}_c(\nu'|\nu) &\leq \int_{X_1 \times X_2} c(x, p) \nu(dx) = \int_{X_1} c_1(x_1, p_1(x_1, \cdot)) \nu_1(dx_1) + \int_{X_1 \times X_2} c_2(x_2, p_2(x, \cdot)) \nu(dx) \\ &\leq \mathcal{T}_{c_1}(\nu'_1|\nu_1) + \varepsilon + \int_{X_1^2 \times X_2} c_2(x_2, q_2^{x_1, y_1}(x_2, \cdot)) p_1(x_1, dy_1) \nu(dx) \\ &= \mathcal{T}_{c_1}(\nu'_1|\nu_1) + \varepsilon + \int_{X_1^2} \left( \int_{X_2} c_2(x_2, q_2^{x_1, y_1}(x_2, \cdot)) \nu_2(x_1, dx_2) \right) p_1(x_1, dy_1) \nu_1(dx_1) \\ &\leq \mathcal{T}_{c_1}(\nu'_1|\nu_1) + \varepsilon + \int_{X_1^2} (\mathcal{T}_{c_2}(\nu'_2(y_1, \cdot)|\nu_2(x_1, \cdot)) + \varepsilon) p_1(x_1, dy_1) \nu_1(dx_1), \end{aligned}$$

where the last two inequalities follow from (5.15) and (5.16) respectively. The expected result follows and the proof of the lemma is complete.  $\square$

*Proof of Theorem 5.11.* By induction, it is enough to consider the case  $n = 2$ . Given  $\nu, \nu' \in \mathcal{P}_\gamma(X_1 \times X_2)$ , thanks to Lemma 5.13, for all  $\varepsilon > 0$ , there exists a kernel  $p_1^\varepsilon$  such that

$$\mathcal{T}_c(\nu'|\nu) \leq \mathcal{T}_{c_1}(\nu'_1|\nu_1) + \int_{X_1 \times X_1} \mathcal{T}_{c_2}(\nu'_2(y_1, \cdot)|\nu_2(x_1, \cdot)) p_1^\varepsilon(x_1, dy_1) \nu_1(dx_1) + 2\varepsilon,$$

where  $\nu, \nu'_1, \nu_2, \nu'_2$  are defined in Lemma 5.13. Applying the transport-entropy inequalities that hold for  $\mu_1$  and  $\mu_2$ , we get

$$\begin{aligned} \mathcal{T}_c(\nu'|\nu) &\leq a_1^{(1)} H(\nu'_1|\mu_1) + a_2^{(1)} H(\nu_1|\mu_1) \\ &\quad + \int_{X_1 \times X_1} \left[ a_1^{(2)} H(\nu'_2(y_1, \cdot)|\mu_2) + a_2^{(2)} H(\nu_2(x_1, \cdot)|\mu_2) \right] p_1^\varepsilon(x_1, dy_1) \nu_1(dx_1) + 2\varepsilon \\ &\leq a_1 \left[ H(\nu'_1|\mu_1) + \int_{X_1} H(\nu'_2(y_1, \cdot)|\mu_2) \nu'_1(dy_1) \right] \\ &\quad + a_2 \left[ H(\nu_1|\mu_1) + \int_{X_1} H(\nu_2(x_1, \cdot)|\mu_2) \nu_1(dx_1) \right] + 2\varepsilon \\ &= a_1 H(\nu'|\mu) + a_2 H(\nu|\mu) + 2\varepsilon, \end{aligned}$$

where we used that  $\int_{X_1} p_1^\varepsilon(x_1, dx'_1) = 1$ ,  $\int_{X_1} p_1^\varepsilon(x_1, \cdot) \nu_1(dx_1) = \nu'_1(\cdot)$  and the chain rule formula for the entropy (recall that  $a_1 := \max(a_1^{(1)}, a_1^{(2)})$  and  $a_2 := \max(a_2^{(1)}, a_2^{(2)})$ ). Letting  $\varepsilon$  go to zero completes the proof of the theorem.  $\square$

**Remark 5.17.** *Alternatively, following [31], one could give a proof based on the dual characterization of Proposition 5.5.*

**5.3. Dimension-free concentration.** In this section, extending [9], we characterize the transport-entropy inequality  $\mathbf{T}_c(a_1, a_2)$  in terms of a dimension-free concentration property. We recall first and introduce some notations.

Let  $\gamma: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be a lower-semicontinuous function satisfying (2.1) and  $c: X \times \mathcal{P}_\gamma(X) \rightarrow [0, \infty)$  such that  $c(x, \delta_x) = 0$  for all  $x \in X$ . Recall from Corollary 5.12 that for all integers  $n \geq 1$ ,

$$c^n(x, p) := \sum_{i=1}^n c(x_i, p_i), \quad x = (x_1, \dots, x_n) \in X^n, \quad p \in \mathcal{P}_r(X^n),$$

where  $p_i$  denotes the  $i$ -th marginal distribution of  $p$ . For all  $\varphi \in \Phi_\gamma(X^n)$ , define as before

$$R_{c^n} \varphi(x) = \inf_{p \in \mathcal{P}_\gamma(X^n)} \left\{ \int \varphi dp + c^n(x, p) \right\}, \quad x \in X^n.$$

Finally for all Borel sets  $A \subset X^n$ , let

$$c_A^n(x) := \inf_{p \in \mathcal{P}_\gamma(X^n): p(A)=1} c^n(x, p), \quad x \in X^n,$$

and, for  $t \geq 0$ ,

$$A_t^n := \{x \in X^n : c_A^n(x) \leq t\}.$$

We are now in a position to state our theorem.

**Theorem 5.18.** *Let  $\gamma: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be a lower-semicontinuous function satisfying (2.1) and  $c: X \times \mathcal{P}_\gamma(X) \rightarrow [0, \infty)$  such that  $c(x, \delta_x) = 0$  for all  $x \in X$  and which satisfies Condition (C), (C') or (C''). For  $\mu \in \mathcal{P}_\gamma(X)$  and  $a_1, a_2 > 0$ , the following are equivalent:*

- (i)  $\mu$  satisfies  $\mathbf{T}_c(a_1, a_2)$ ;
- (ii) there exists a numerical constant  $K$  such that for all integers  $n \geq 1$ , for all Borel sets  $A \subset X^n$ , it holds

$$(5.19) \quad \mu^n(X^n \setminus A_t^n)^{a_2} \mu^n(A)^{a_1} \leq K e^{-t} \quad \forall t \geq 0.$$

- (iii) there exists a numerical constant  $K$  such that for all integers  $n \geq 1$ , for all non-negative  $\varphi \in \Phi_\gamma(X^n)$ , it holds

$$\mu^n(R_{c^n} \varphi > u)^{a_2} \mu^n(\varphi \leq v)^{a_1} \leq K e^{-u+v} \quad \forall u, v \in \mathbb{R}.$$

**Remark 5.20.**

- (1) Note that when  $c(x, p) = \int \omega(x, y) p(dy)$ , for some measurable  $\omega: X \times X \rightarrow [0, \infty)$ , the enlargement  $A_t^n$  of some set  $A \subset X$  reduces to

$$A_t^n = \{x \in X^n; \exists y \in A \text{ s.t. } \sum_{i=1}^n \omega(x_i, y_i) \leq t\}.$$

In particular, when  $X = \mathbb{R}^m$  and  $\omega(x, y) = \|x - y\|^r$ ,  $r \geq 2$ , where  $\|\cdot\|$  is a given norm on  $\mathbb{R}^m$ , then denoting by

$$(5.21) \quad B_r^n = \{x \in (\mathbb{R}^m)^n; \sum_{i=1}^n \|x_i\|^r \leq 1\},$$

then it holds

$$A_t^n = A + t^{1/r} B_r^n.$$

- (2) When  $c(x, p) = \|x - \int y p(dy)\|^r$ ,  $x \in \mathbb{R}^m, p \in \mathcal{P}_1(\mathbb{R}^m)$ , for some norm  $\|\cdot\|$  on  $\mathbb{R}^m$ , then the enlargement of a set  $A \subset (\mathbb{R}^m)^n$  reduces to

$$A_t^n = \overline{\text{conv}}(A) + t^{1/r} B_r^n,$$



denoting by  $\overline{\text{conv}}(A)$  the closed convex hull of  $A$ . Indeed, denoting by  $\|\cdot\|_r^n$  the norm defined on  $(\mathbb{R}^m)^n$  by  $\|x\|_r^n = (\sum_{i=1}^n \|x_i\|_r^r)^{1/r}$ , then, for all  $x \in (\mathbb{R}^m)^n$ , it holds  $c_A(x) = \inf_{y \in C} \{\|x - y\|_r^n\} = \inf_{y \in \overline{C}} \{\|x - y\|_r^n\}$ , with  $C = \{\int y p(dy); p \in \mathcal{P}_1(A)\}$ . It is well known that  $\overline{C} = \overline{\text{conv}}(A)$ , which proves the claim.

- (3) The implications (i)  $\Rightarrow$  (ii) and (ii)  $\Rightarrow$  (iii) are in fact true solely under the assumptions  $c(x, \delta_x) = 0$  for all  $x \in X$  and  $(C_3)$  ( $p \mapsto c(x, p)$  is convex), as the proof indicates.

*Proof.* First we prove that (i) implies (ii). Since  $\mu$  satisfies  $\mathbf{T}_c(a_1, a_2)$ , by the tensorisation property, for all positive integers  $n$ , it holds

$$\mathcal{T}_{c^n}(\nu_1 | \nu_2) \leq a_1 H(\nu_1 | \mu^n) + a_2 H(\nu_2 | \mu^n),$$

for all  $\nu_1, \nu_2 \in \mathcal{P}_\gamma(X^n)$ . Let  $A \subset X^n$  be a Borel set and define  $\nu_1(dx) = \frac{\mathbf{1}_A(x)}{\mu^n(A)} \mu^n(dx)$  and  $\nu_2(dx) = \frac{\mathbf{1}_B(x)}{\mu^n(B)} \mu^n(dx)$ , where  $B = X^n \setminus A_t^n$ , for some  $t > 0$ . Then  $H(\nu_1 | \mu^n) = -\log \mu^n(A)$  and  $H(\nu_2 | \mu^n) = -\log \mu^n(B)$ . Furthermore, if  $\pi \in \Pi(\nu_2, \nu_1)$  with disintegration kernel  $(p_x)_{x \in X^n}$ , then for  $\nu_2$  almost all  $x \in X^n$ ,  $p_x(A) = 1$ . Therefore,

$$\int c(x, p_x) \nu_2(dx) \geq \int c_A^n(x) \frac{\mathbf{1}_B(x)}{\mu^n(B)} \mu^n(dx) \geq t,$$

where the last inequality comes from the fact that  $c_A^n(x) > t$  for all  $x \in B = \{x \in X^n : c_A^n(x) > t\}$ . Taking the infimum over all  $\pi \in \Pi(\nu_2, \nu_1)$  finally yields

$$t \leq \mathcal{T}_{c^n}(\nu_1 | \mu_2) \leq -a_1 \log(\mu^n(A)) - a_2 \log \mu^n(X^n \setminus A_t^n),$$

which proves (ii).

Now we prove that (ii) implies (iii). Fix  $n \geq 1$ ,  $m \in \mathbb{R}$ ,  $t \geq 0$  and a non-negative  $\varphi \in \Phi_\gamma(X^n)$ . We will prove that  $\{R_{c^n} \varphi > m + t\} \subset \{c_A^n > t\}$  with  $A := \{\varphi \leq m\}$ . To that aim consider  $x \in \{R_{c^n} \varphi > m + t\}$ . Then, for all  $p \in \mathcal{P}_\gamma(X^n)$  with  $p(A) = 1$ , we have  $\int \varphi dp \leq m$  so that, by definition of  $R_{c^n}$ , it holds

$$m + t < \int \varphi dp + c^n(x, p) \leq m + c^n(x, p).$$

Hence, taking the infimum over all  $p$  with  $p(A) = 1$  leads to  $c_A^n(x) > t$ , which is the desired result. Point (iii) then immediately follows applying Point (ii) to  $A$ .

Finally we prove that (iii) implies (i), following [11]. Fix  $\varepsilon \in (0, 1)$ . Given  $f \in \Phi_\gamma(X)$ , non-negative, let  $\varphi(x) = f(x_1) + f(x_2) + \dots + f(x_n)$ ,  $x \in X^n$ . Then,  $\varphi \in \Phi_\gamma(X^n)$  is also non-negative and  $R_{c^n} \varphi(x) = \sum_{i=1}^n R_c f(x_i)$ , so that, using the product structure of  $\mu^n$ ,

$$(5.22) \quad \left( \int e^{\frac{R_c f}{(1+\varepsilon)a_2}} d\mu \right)^{a_2} \left( \int e^{-\frac{f}{(1-\varepsilon)a_1}} d\mu \right)^{a_1} = \left( \int e^{\frac{R_{c^n} \varphi}{(1+\varepsilon)a_2}} d\mu^n \right)^{a_2/n} \left( \int e^{-\frac{\varphi}{(1-\varepsilon)a_1}} d\mu^n \right)^{a_1/n}.$$

Our aim is to prove that the right hand side, to the power  $n$ , is bounded. Thanks to Point (iii), for any  $v \in \mathbb{R}$  it holds

$$\begin{aligned} \int e^{\frac{R_{c^n} \varphi}{(1+\varepsilon)a_2}} d\mu^n &= 1 + \int_0^\infty e^u \mu^n \left( \frac{R_{c^n} \varphi}{(1+\varepsilon)a_2} > u \right) du \\ &\leq 1 + \mu^n \left( \frac{\varphi}{(1-\varepsilon)a_1} \leq v \right)^{-\frac{a_1}{a_2}} K^{\frac{1}{a_2}} e^{\frac{(1-\varepsilon)a_1 v}{a_2}} \int_0^\infty e^{-\varepsilon u} du \\ &= 1 + \frac{1}{\varepsilon} \mu^n \left( \frac{\varphi}{(1-\varepsilon)a_1} \leq v \right)^{-\frac{a_1}{a_2}} K^{\frac{1}{a_2}} e^{\frac{(1-\varepsilon)a_1 v}{a_2}}. \end{aligned}$$

In particular, for all  $v \in \mathbb{R}$ ,

$$\left(-1 + \int e^{\frac{R_c n \varphi}{(1+\varepsilon)a_2}} d\mu^n\right)^{\frac{a_2}{a_1}} e^{-v} \mu^n\left(\frac{\varphi}{(1-\varepsilon)a_1} \leq v\right) \leq K^{\frac{1}{a_1}} \frac{e^{-\varepsilon v}}{\varepsilon^{\frac{a_2}{a_1}}}.$$

Since  $\int e^{-\frac{\varphi}{(1-\varepsilon)a_1}} d\mu^n = \int_0^\infty e^{-v} \mu^n\left(\frac{\varphi}{(1-\varepsilon)a_1} \leq v\right) dv$ , integrating the latter implies that

$$\left(-1 + \int e^{\frac{R_c n \varphi}{(1+\varepsilon)a_2}} d\mu^n\right)^{\frac{a_2}{a_1}} \int e^{-\frac{\varphi}{(1-\varepsilon)a_1}} d\mu^n \leq \frac{K^{\frac{1}{a_1}}}{\varepsilon^{1+\frac{a_2}{a_1}}}.$$

This in turn implies, by simple algebra that

$$\begin{aligned} & \left(\int e^{\frac{R_c n \varphi}{(1+\varepsilon)a_2}} d\mu^n\right)^{a_2} \left(\int e^{-\frac{\varphi}{(1-\varepsilon)a_1}} d\mu^n\right)^{a_1} \\ & \leq \left(1 + \frac{1}{\varepsilon} \left(\varepsilon \int e^{-\frac{\varphi}{(1-\varepsilon)a_1}} d\mu^n\right)^{-\frac{a_1}{a_2}}\right)^{a_2} \left(\int e^{-\frac{\varphi}{(1-\varepsilon)a_1}} d\mu^n\right)^{a_1} \\ & = \left(\left(\int e^{-\frac{\varphi}{(1-\varepsilon)a_1}} d\mu^n\right)^{\frac{a_1}{a_2}} + \frac{1}{\varepsilon^{1+\frac{a_1}{a_2}}}\right)^{a_2} \leq \left(1 + \frac{1}{\varepsilon^{1+\frac{a_1}{a_2}}}\right)^{a_2}, \end{aligned}$$

where in the last line we used that  $\varphi$  is a non-negative function.

Plugging this bound into (5.22) leads, in the limit  $n \rightarrow \infty$ , to

$$\left(\int e^{\frac{R_c f}{(1+\varepsilon)a_2}} d\mu\right)^{a_2} \left(\int e^{-\frac{f}{(1-\varepsilon)a_1}} d\mu\right)^{a_1} \leq 1.$$

Taking  $\varepsilon$  to 0 gives  $\mathbf{T}_c(a_1, a_2)$ , thanks to Proposition 5.5.  $\square$

Concentration of measure inequalities are usually stated for enlargements of sets of measure bigger than  $1/2$  (see [18]). In what follows we connect (5.19) to the usual definition for some families of cost functionals.

**Lemma 5.23.** *Consider a cost function  $c$  of the form*

$$c(x, p) = \int \gamma(d(x, y)) p(dy), \quad x \in X, \quad p \in \mathcal{P}_\gamma(X)$$

with  $\gamma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  an increasing convex function such that  $\gamma(0) = \gamma'(0) = 0$  and suppose that  $\gamma$  satisfies (2.1). Suppose that, for a given  $n \in \mathbb{N}^*$ , a probability measure  $\mu$  on  $X$  satisfies, for some constants  $a > 0, b \geq 1$ , the following concentration property :

$$(5.24) \quad \mu^n(X^n \setminus A_t^n) \leq b e^{-t/a}, \quad \forall t \geq 0,$$

for all  $A \subset X^n$  such that  $\mu^n(A) \geq 1/2$ .

Then  $\mu$  satisfies the following property : for all  $s \in (0, 1)$  and for all  $A \subset X^n$ ,

$$(5.25) \quad \mu^n(X^n \setminus A_t^n)^{1/(1-s)^{r-1}} \mu^n(A)^{1/s^{r-1}} \leq b e^{-t/a}, \quad \forall t \geq 0,$$

where the exponent  $r$  is defined by  $r = \sup_{x>0} x \gamma'_+(x) / \gamma(x) \in (1, \infty)$  (here  $\gamma'_+$  stands for the right derivative).

Conversely, if the concentration property (5.25) holds, then one has (by optimizing over all  $s \in (0, 1)$ ), for all  $A \subset X^n$  such that  $\mu^n(A) \geq 1/2$ , for all  $t > \max(a \log(2b), 0)$ ,

$$\mu^n(X^n \setminus A_t^n) \leq \inf_{s \in (0, 1)} \left( b^{(1-s)^{r-1}} 2^{\frac{(1-s)^{r-1}}{s^{r-1}}} e^{-t(1-s)^{r-1}/a} \right) = b e^{-t(1-\varepsilon(t))^r/a},$$

with  $\varepsilon(t) = \left(\frac{\log 2}{\frac{t}{a} - \log b}\right)^{1/r}$ .

*Proof.* The fact that  $1 < r < \infty$  follows from (2.1) and the convexity inequalities  $\gamma(2x) \geq \gamma(x) + x\gamma'(x)$  and  $\gamma(x)/x < \gamma'(x)$ ,  $x > 0$ .

To clarify the notations, we will omit some of the dependencies in  $n$  in this proof. The fact that (5.24) implies (5.25) is a consequence of the following sets inclusions (that are justified at the end of the proof):

$$(a) \quad A \subset X^n \setminus ((X^n \setminus A_u)_u), \quad \forall u \geq 0,$$

and for all  $s \in (0, 1)$ ,

$$(b) \quad (A_u)_v \subset A_{(u^{1/r} + v^{1/r})^r} \subset A_{\frac{u}{s^{r-1}} + \frac{v}{(1-s)^{r-1}}}, \quad \forall u, v \geq 0.$$

The last inclusion above follows from the identity,

$$(5.26) \quad \left(u^{1/r} + v^{1/r}\right)^r = \inf_{s \in (0,1)} \left\{ \frac{u}{s^{r-1}} + \frac{v}{(1-s)^{r-1}} \right\}.$$

Let  $t \geq 0$ ,  $s \in (0, 1)$  and  $A \subset X^n$  and let us consider the set  $B = A_{s^{r-1}t}$ .

If  $\mu(B) \geq 1/2$  then by applying first (b) for  $u = s^{r-1}t$  and  $v = (1-s)^{r-1}t$  and then the concentration property (5.24), we get

$$\mu(X^n \setminus A_t) \leq \mu\left(X^n \setminus B_{(1-s)^{r-1}t}\right) \leq be^{-(1-s)^{r-1}t/a}.$$

If  $\mu(B) < 1/2$  then  $\mu(X^n \setminus B) \geq 1/2$ . Therefore by applying first (a) for  $u = s^{r-1}t$  and then the concentration property (5.24), we get

$$\mu(A) \leq \mu\left(X^n \setminus ((X^n \setminus B)_{s^{r-1}t})\right) \leq be^{-s^{r-1}t/a}.$$

As a consequence in any case the concentration property (5.25) holds.

Now let us justify the inclusion properties (a) and (b).

To prove (a) let us show that  $A \cap (X^n \setminus A_u)_u = \emptyset$ . Suppose on the contrary that there is some  $x \in A \cap (X^n \setminus A_u)_u$ , then there is some  $y \in X^n \setminus A_u$  such that  $\sum_{i=1}^n \gamma(d(x_i, y_i)) \leq u$ . But, since  $y \in X^n \setminus A_u$ , it holds  $\sum_{i=1}^n \gamma(d(y_i, z_i)) > u$  for all  $z \in A$ . In particular, taking  $z = x$ , one gets a contradiction.

Finally, let us show (b). According to e.g. [13, Lemma 4.7], the function  $x \mapsto \gamma^{1/r}(x)$  is subadditive. It follows easily that  $(x, y) \mapsto (\sum_{i=1}^n \gamma(d(x_i, y_i)))^{1/r}$  defines a distance on  $X^n$ . Point (b) then follows immediately from the triangle inequality.  $\square$

For the next corollaries, recall the notations introduced in Remark 5.20.

**Corollary 5.27.** *Let  $r \geq 2$  and consider the cost  $c(x, p) = \int \|x - y\|^r p(dy)$ ,  $x \in \mathbb{R}^m$ ,  $p \in \mathcal{P}_1(\mathbb{R}^m)$ , where  $\|\cdot\|$  is a norm on  $\mathbb{R}^m$ . For a probability  $\mu \in \mathcal{P}_r(\mathbb{R}^m)$ , the following propositions are equivalent :*

(1) *There exist  $a_1, b_1 > 0$  such that,  $\forall n \in \mathbb{N}^*$ ,*

$$\mu^n(A + t^{1/r} B_r^n) \geq 1 - b_1 e^{-t/a_1}, \quad \forall t \geq 0,$$

*for all sets  $A$  such that  $\mu^n(A) \geq 1/2$ .*

(2) *There exist  $a_2, b_2 > 0$  such that,  $\forall n \in \mathbb{N}^*$ ,*

$$\mu^n(f > \text{med}(f) + r) \leq b_2 e^{-t/a_2}, \quad \forall t \geq 0,$$

*for all  $f : (\mathbb{R}^m)^n \rightarrow \mathbb{R}$  which are 1-Lipschitz with respect to the norm  $\|\cdot\|_r^n$  defined on  $(\mathbb{R}^m)^n$  by*

$$\|x\|_r^n = \left( \sum_{i=1}^n \|x_i\|^r \right)^{1/r}, \quad x \in (\mathbb{R}^m)^n.$$

(3) There exist  $a_3, b_3 > 0$  such that,  $\forall n \in \mathbb{N}^*$ ,  $\forall s \in (0, 1)$ , and  $\forall A \subset (\mathbb{R}^m)^n$ ,

$$\mu^n((\mathbb{R}^m)^n \setminus A_t^n)^{1/(1-s)^{r-1}} \mu^n(A)^{1/s^{r-1}} \leq b_3 e^{-t/a_3}, \quad \forall t \geq 0,$$

where  $A_t^n = \{x \in (\mathbb{R}^m)^n; c_A^n(x) \leq t\} = A + t^{1/r} B_r^n$ .

(4)  $\exists a_4 > 0$  such that  $\forall s \in (0, 1)$ ,  $\mu$  satisfies

$$\mathbf{T}_r(a_4/s^{r-1}, a_4/(1-s)^{r-1}).$$

(5)  $\exists a_5 > 0$  such that  $\mu$  satisfies  $\mathbf{T}_r^+(a_5)$  (which is equivalent to  $\mathbf{T}_r^-(a_5)$  for that cost).

Moreover (1)  $\Leftrightarrow$  (2) with  $a_2 = a_1$  and  $b_2 = b_1$ , (3)  $\Rightarrow$  (4) with  $a_4 = a_3$ , (4)  $\Rightarrow$  (3) with  $a_3 = a_4$  and  $b_3 = 1$ , (4)  $\Leftrightarrow$  (5) with  $a_4 = a_5$ , (1)  $\Rightarrow$  (3) with  $a_3 = a_1$  and  $b_3 = b_1$ , (3)  $\Rightarrow$  (1) with  $b_1 = b_3^{(1-s)^{r-1}} 2^{\frac{(1-s)^{r-1}}{s^{r-1}}}$  and  $a_1 = \frac{a_3}{(1-s)^{r-1}}$  for any  $s \in (0, 1)$ .

Note that this result is not as general as possible; see [9, Theorem 1.3] for a similar statement involving more general cost functions.

*Proof.* The equivalence (1)  $\Leftrightarrow$  (2) is very classical (see e.g [18, Proposition 1.3]).

The implications (1)  $\Rightarrow$  (3) and (3)  $\Rightarrow$  (1) are given in Lemma 5.23. (3)  $\Rightarrow$  (4) and (4)  $\Rightarrow$  (3) are consequences of Theorem 5.18.

If the property (4) holds, then for all  $\nu_1 \in \mathcal{P}_r$ ,

$$\mathcal{T}_r(\nu_1, \mu) = \mathcal{T}_c(\nu_1 | \mu) \leq \frac{a_4}{s^{r-1}} H(\nu_1 | \mu) \quad \forall s \in (0, 1).$$

As  $s$  goes to 1, we get (5),  $\mu$  satisfies  $\mathbf{T}_r^+(a_4)$  or equivalently  $\mathbf{T}_r^-(a_4)$ .

Conversely assume that (5) holds. By the triangular inequality, we get for all  $\nu_1, \nu_2 \in \mathcal{P}_r$ ,

$$\begin{aligned} \mathcal{T}_c(\nu_1 | \nu_2) &= \mathcal{T}_r(\nu_1, \nu_2) \leq \left( \mathcal{T}_r(\nu_1, \mu)^{1/r} + \mathcal{T}_r(\mu, \nu_2)^{1/r} \right)^r \\ &\leq \left( (a_5 H(\nu_1 | \mu))^{1/r} + (a_5 H(\nu_2 | \mu))^{1/r} \right)^r. \end{aligned}$$

The property (4) with  $a_4 = a_5$  then follows from the identity (5.26).  $\square$

**Corollary 5.28.** Let  $r \geq 2$  and consider the cost  $c(x, p) = \|x - \int y p(dy)\|^r$ ,  $x \in \mathbb{R}^m$ ,  $p \in \mathcal{P}_1(\mathbb{R}^m)$ . For a probability  $\mu \in \mathcal{P}_1(\mathbb{R}^m)$ , the following propositions are equivalent :

(1) There exist  $a_1, b_1 > 0$  such that,  $\forall n \in \mathbb{N}^*$ ,

$$\mu^n(A + t^{1/r} B_r^n) \geq 1 - b_1 e^{-t/a_1}, \quad \forall t \geq 0,$$

for any set  $A$  which is either convex or the complement of a convex set and such that  $\mu^n(A) \geq 1/2$ .

(2) There exist  $a_2, b_2 > 0$  such that,  $\forall n \in \mathbb{N}^*$ ,

$$\mu^n(f > \text{med}(f) + r) \leq b_2 e^{-t/a_2}, \quad \forall t \geq 0,$$

for all  $f : (\mathbb{R}^m)^n \rightarrow \mathbb{R}$  which is either convex or concave and 1-Lipschitz with respect to the norm  $\|\cdot\|_r^n$  defined on  $(\mathbb{R}^m)^n$  by

$$\|x\|_r^n = \left( \sum_{i=1}^n \|x_i\|^r \right)^{1/r}, \quad x \in (\mathbb{R}^m)^n.$$

(3) There exist  $a_3, b_3 > 0$  such that,  $\forall n \in \mathbb{N}^*$ ,  $\forall s \in (0, 1)$ , and  $\forall A \subset (\mathbb{R}^m)^n$ ,

$$\mu^n((\mathbb{R}^m)^n \setminus A_t^n)^{1/(1-s)^{r-1}} \mu^n(A)^{1/s^{r-1}} \leq b_3 e^{-t/a_3}, \quad \forall t \geq 0,$$

where  $A_t^n = \{x \in (\mathbb{R}^m)^n; c_A^n(x) \leq t\} = \overline{\text{conv}} A + t^{1/r} B_r^n$ .

(4) There exists  $a_4 > 0$  such that  $\forall s \in (0, 1)$ ,  $\mu$  satisfies  $\overline{\mathbf{T}}_r(a_4/s^{r-1}, a_4/(1-s)^{r-1})$ .

(5) There exists  $a_5 > 0$  such that  $\mu$  satisfies  $\overline{\mathbf{T}}_r^+(a_5)$  and  $\mu$  satisfies  $\overline{\mathbf{T}}_r^-(a_5)$ .

Moreover (1)  $\Leftrightarrow$  (2) with  $a_2 = a_1$  and  $b_2 = b_1$ , (3)  $\Rightarrow$  (4) with  $a_4 = a_3$ , (4)  $\Rightarrow$  (3) with  $a_3 = a_4$  and  $b_3 = 1$ , (4)  $\Leftrightarrow$  (5) with  $a_4 = a_5$ , (1)  $\Rightarrow$  (3) with  $a_3 = a_1$  and  $b_3 = b_1$ , (3)  $\Rightarrow$  (1) with  $b_1 = b_3^{(1-s)^{r-1}} 2^{\frac{(1-s)^{r-1}}{s^{r-1}}}$  and  $a_1 = \frac{a_3}{(1-s)^{r-1}}$  for any  $s \in (0, 1)$ .

*Proof.* Adapting [18, Proposition 1.3], one sees easily that (1)  $\Leftrightarrow$  (2), and, according to Theorem 5.18, (3)  $\Leftrightarrow$  (4).

Let us show that (3) implies (1). Let  $A$  be a convex subset. As in Lemma 5.23, if  $\mu^n(A) \geq 1/2$ , then, by applying (3) to  $A$  and since  $A_t = \overline{A} + t^{1/r} B_r^n$ , we get (1) for convex sets with  $b_1 = b_3^{(1-s)^{r-1}} 2^{\frac{(1-s)^{r-1}}{s^{r-1}}}$  and  $a_1 = \frac{a_3}{(1-s)^{r-1}}$  for  $s \in (0, 1)$ . Let  $D = (\mathbb{R}^m)^n \setminus A$  and assume that  $\mu(D) \geq 1/2$ . For all  $t > 0$ , the set  $C = (\mathbb{R}^m)^n \setminus (D + t^{1/r} B_r^n)$  is convex and satisfies for all  $t' < t$ ,

$$C_{t'} = (\overline{C} + t'^{1/r} B_r^n) \subset (\mathbb{R}^m)^n \setminus D.$$

Since  $\mu^n(D) \geq 1/2$ , it follows that  $\mu^n((\mathbb{R}^m)^n \setminus C_{t'}) \geq 1/2$ . As a consequence, applying (3) to the set  $C$ , we obtain for all  $t > t' > 0$ , for all  $s \in (0, 1)$ ,

$$\mu^n((\mathbb{R}^m)^n \setminus (D + t^{1/r} B_r^n)) = \mu^n(C) \leq b_3^{s^{r-1}} 2^{\frac{s^{r-1}}{(1-s)^{r-1}}} e^{-\frac{s^{r-1} t'}{a_3}}.$$

As  $t'$  goes to  $t$ , this implies the concentration property (1) for complement of convex sets.

We adapt the proof of Lemma 5.23 to get (1)  $\Rightarrow$  (3). The property (a) is replaced by the following, for all subset  $A$ ,

$$(a') \quad A \subset \overline{\text{conv}} A \subset (\mathbb{R}^m)^n \setminus [(X \setminus A_u) + u^{1/r} B_r^n], \quad u \geq 0.$$

Since  $A_u = \overline{\text{conv}} A + u^{1/r} B_r^n$ , this property (a') is a simple consequence of the property (a) applied to the set  $\overline{\text{conv}} A$ . For the same reason, the property (b) still holds. Then following the proof of Lemma 5.23, by using (a') and (b), by considering the set  $B = A_{s^{r-1}t}$ ,  $s \in (0, 1)$ , and applying the concentration property (1) to the convex set  $B$  or to its complement  $(\mathbb{R}^m)^n \setminus B$ , we get (1)  $\Rightarrow$  (3) with  $a_3 = a_1$  and  $b_3 = b_1$ .

The equivalence between (3) and (4) is a consequence of Theorem 5.18.

If the property (4) holds, then for all  $\nu_1, \nu_2 \in \mathcal{P}_r$ ,

$$\overline{\mathcal{T}}_r(\nu_1|\mu) \leq \frac{a_4}{s^{r-1}} H(\nu_1|\mu), \quad \text{and} \quad \overline{\mathcal{T}}_r(\mu|\nu_2) \leq \frac{a_4}{(1-s)^{r-1}} H(\nu_2|\mu), \quad \forall s \in (0, 1).$$

As  $s$  goes to 1 or to 0, we get (5) – that  $\mu$  satisfies  $\overline{\mathbf{T}}_r^+(a_4)$  and  $\overline{\mathbf{T}}_r^-(a_4)$ .

Conversely assume that (5) holds, then (4) follows with  $a_4 = a_5$  by the following triangular inequality, for all  $\nu_1, \nu_2 \in \mathcal{P}_r$ ,

$$\overline{\mathcal{T}}_r(\nu_1|\nu_2)^{1/r} \leq \overline{\mathcal{T}}_r(\nu_1|\mu)^{1/r} + \overline{\mathcal{T}}_r(\mu|\nu_2)^{1/r}.$$

□

## 6. WEAK TRANSPORT-ENTROPY AND LOG-SOBOLEV TYPE INEQUALITIES

In this section, our aim is to give some explicit links between the weak transport-entropy inequalities introduced in Definition 5.1 and functional inequalities of log-Sobolev type. Except for the first result below, we are not able to deal with general costs. Hence (except for Section 6.1), we restrict to the specific case (already of interest) of  $\overline{\mathcal{T}}_\theta$  (introduced in Section 3). Furthermore, to avoid technicalities, we may restrict to the particular choice  $\theta(x) = \|x\|^2$  (for some norm on  $\mathbb{R}^m$ ), even if most of the results below could be extended to more general convex functions (at the price of denser statements and more technical proofs).

**6.1. Transport-entropy and  $(\tau)$ -log-Sobolev inequalities.** In this section we generalize the notion of  $(\tau)$ -log-Sobolev inequality introduced in [12] (see also [13]) and give some connection with weak transport-entropy inequalities.

We need some notations. Given  $\lambda > 0$  and  $\varphi \in \Phi_\gamma(X)$ , define

$$R_c^\lambda \varphi(x) := \inf_{p \in \mathcal{P}_\gamma(X)} \left\{ \int \varphi(y) p(dy) + \lambda c(x, p) \right\}, \quad x \in X.$$

Observe that  $R_c^1 = R_c$ , where  $R_c$  is defined in Theorem 3.5. Following [12], we introduce the  $(\tau)$ -log-Sobolev inequality as follows. We recall that for any non-negative function  $g$ , one denotes  $\text{Ent}_\mu(g) = \int g \log \left( \frac{g}{\int g d\mu} \right) d\mu$ .

**Definition 6.1** ( $(\tau) - \text{LSI}_c(\lambda, C)$ ). *Let  $\gamma: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be a lower-semicontinuous function satisfying (2.1),  $c: X \times \mathcal{P}_\gamma(X) \rightarrow [0, \infty)$  and  $C \in (0, \infty)$ . Then  $\mu \in \mathcal{P}_\gamma(X)$  is said to satisfy the  $(\tau)$ -log-Sobolev inequality with constant  $C, \lambda$  and cost  $c$  (or in short  $(\tau) - \text{LSI}_c(\lambda, C)$ ) if, for all  $f$  with  $\int f e^f d\mu < \infty$ , it holds*

$$(6.2) \quad \text{Ent}_\mu(e^f) \leq C \int (f - R_c^\lambda f) e^f d\mu.$$

The following result extends [12, Theorem 2.1].

**Proposition 6.3.** *Let  $\gamma: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be a lower-semicontinuous function satisfying (2.1) and  $c: X \times \mathcal{P}_\gamma(X) \rightarrow [0, \infty)$  be a cost function. If  $\mu \in \mathcal{P}_\gamma(X)$  satisfies  $\mathbf{T}_c^-(b)$ , then it satisfies  $(\tau) - \text{LSI}_c(\lambda, \frac{1}{1-\lambda b})$  for all  $\lambda \in (0, 1/b)$ .*

**Remark 6.4.** *In  $\mathbb{R}^n$  [12], and more generally in metric spaces [13], if one considers the usual transport cost  $\mathcal{T}_2$  (with cost  $\omega(x, y) = d(x, y)^2$  in Item (1) of Section 3), it is proved that the corresponding  $\mathbf{T}_2^-(b)$  is actually equivalent to some  $(\tau)$ -log-Sobolev inequality. In order to get such a result in the setting of the present paper one would need to develop a general Hamilton-Jacobi theory which is not available at present (see [32] for some developments). This is mainly the reason why, in the next sections, we restrict ourselves to the specific case of the “bar” cost.*

*Proof.* Fix a function  $f: X \rightarrow \mathbb{R}$  with  $\int f e^f d\mu < \infty$ ,  $\lambda \in (0, 1/C)$  and define  $d\nu_f = \frac{e^f}{\int e^f d\mu} d\mu$ . One has

$$H(\nu_f | \mu) = \int \log \left( \frac{e^f}{\int e^f d\mu} \right) \frac{e^f}{\int e^f d\mu} d\mu = \int f d\nu_f - \log \int e^f d\mu \leq \int f d\nu_f - \int f d\mu,$$

where the last inequality comes from Jensen’s inequality. Consequently, if  $\pi(dx dy) = \nu_f(dx) p_x(dy)$  is a probability measure on  $X \times X$  with first marginal  $\nu_f$  and second marginal  $\mu$ ,

$$H(\nu_f | \mu) \leq \iint (f(x) - f(y)) \pi(dx dy) = \int \left( \int (f(x) - f(y)) p_x(dy) \right) \nu_f(dx).$$

It follows from the definition of  $R_c^\lambda$  that  $-\int f(y) p_x(dy) \leq -R_c^\lambda f(x) + \lambda c(x, p_x)$  for all  $x \in X$ , so using that  $p_x$  is a probability measure,

$$\int (f(x) - f(y)) p_x(dy) = f(x) - \int f(y) p_x(dy) \leq f(x) - R_c^\lambda f(x) + \lambda c(x, p_x), \quad x \in X.$$

Hence,

$$H(\nu_f | \mu) \leq \int (f(x) - R_c^\lambda f(x)) \nu_f(dx) + \lambda \int c(x, p_x) \nu_f(dx).$$

Optimizing over all  $\pi$  (or equivalently over all  $p_x$ ) with marginals  $\nu_f$  and  $\mu$ , it holds

$$\begin{aligned} H(\nu_f|\mu) &\leq \int \left( f(x) - R_c^\lambda f(x) \right) \nu_f(dx) + \lambda \mathcal{T}_c(\mu|\nu_f) \\ &\leq \frac{1}{\int e^f d\mu} \int \left( f - R_c^\lambda f \right) e^f d\mu + \lambda b H(\nu_f|\mu). \end{aligned}$$

The thesis follows by noticing that  $\left( \int e^f d\mu \right) H(\nu_f|\mu) = \text{Ent}_\mu(e^f)$ .  $\square$

**6.2. Weak transport-entropy inequalities  $\overline{\mathbf{T}}_2^\pm$ .** In this section we give different equivalent forms of  $\overline{\mathbf{T}}_2^\pm$  in terms of the classical log-Sobolev-type inequality of Gross [16] restricted to convex/concave functions, to the  $(\tau)$ -log-Sobolev inequality (6.2) and to the hypercontractivity of the (classical) Hamilton Jacobi semi-group, also restricted to some class of functions.

Throughout this section, we consider the cost  $c(x, p) = \frac{1}{2} \|x - \int y p(dy)\|^2$ ,  $x \in \mathbb{R}^m$ ,  $p \in \mathcal{P}_1(\mathbb{R}^m)$ , where  $\|\cdot\|$  is a norm on  $\mathbb{R}^m$  whose dual norm we denote by  $\|\cdot\|_*$ . We recall that  $\|x\|_* = \max_{y \in \mathbb{R}^m, \|y\|=1} x \cdot y$ . Recall the definition of  $\overline{\mathcal{T}}_2$  from Section 3 and the  $(\tau)$ -log-Sobolev inequality (6.2) defined with such a cost. As usual,  $\|f\|_p := \left( \int |f|^p d\mu \right)^{\frac{1}{p}}$ ,  $p \in \mathbb{R}^*$  (including negative real numbers) and  $\|f\|_0 := \exp\{\int f \log |f| d\mu\}$  whenever this makes sense. Also, given  $\varphi: \mathbb{R}^m \rightarrow \mathbb{R}$ ,  $t > 0$ , let

$$(6.5) \quad Q_t \varphi(x) := \inf_{y \in \mathbb{R}^m} \left\{ \varphi(y) + \frac{1}{2t} \|x - y\|^2 \right\}, \quad P_t \varphi(x) := \sup_{y \in \mathbb{R}^m} \left\{ \varphi(y) - \frac{1}{2t} \|x - y\|^2 \right\}, \quad x \in \mathbb{R}^m.$$

We will make use of the following observation (see Corollary 3.11): for any  $\varphi: \mathbb{R}^m \rightarrow \mathbb{R}$  convex, Lipschitz and bounded from below,  $Q_1 \varphi = R_c \varphi = \inf_{p \in \mathcal{P}_1(X)} \left\{ \int \varphi(y) p(dy) + c(x, p) \right\}$ .

In the result below, we assume that  $\|\cdot\|_*$  is *strictly convex*, i.e. it is such that

$$(6.6) \quad (x \neq y \text{ with } \|x\|_* = \|y\|_* = 1) \Rightarrow \|(1-t)x + ty\|_* < 1.$$

This assumption is made to ensure that the operation  $f \mapsto Q_t f$  transforms a convex function into a  $\mathcal{C}^1$ -smooth convex function (this well known property is recalled in Lemma 6.12 below). The proof could certainly be adapted without this assumption, but we don't want to enter into these technical complications.

**Remark 6.7.** *It is well known that the strict convexity of the dual norm  $\|\cdot\|_*$  is equivalent to the  $\mathcal{C}^1$ -smoothness of the initial norm  $\|\cdot\|$  on  $\mathbb{R}^m \setminus \{0\}$ . These equivalent conditions are fulfilled for instance by the classical  $p$ -norms :  $\|x\|_p = [\sum_{i=1}^m |x_i|^p]^{1/p}$ ,  $x \in \mathbb{R}^m$ , for  $1 < p < +\infty$ .*

**Theorem 6.8.** *Suppose that  $\|\cdot\|_*$  is a strictly convex norm and let  $\mu \in \mathcal{P}_1(\mathbb{R}^m)$ . Then the following are equivalent:*

- (i) *there exists  $b > 0$  such that  $\overline{\mathbf{T}}_2^-(b)$  holds;*
- (ii) *there exists  $\lambda, C > 0$  such that  $(\tau) - \mathbf{LSI}_c(\lambda, C)$  holds;*
- (iii) *there exists  $\rho > 0$  such that for all  $\mathcal{C}^1$ -smooth function  $\varphi: \mathbb{R}^m \rightarrow \mathbb{R}$  convex, Lipschitz and bounded from below, it holds*

$$(6.9) \quad \text{Ent}_\mu(e^\varphi) \leq \frac{1}{2\rho} \int \|\nabla \varphi\|_*^2 e^\varphi d\mu.$$

- (iv) *There exists  $\rho' > 0$  such that for every  $t > 0$ , every  $a \geq 0$  and every  $\varphi: \mathbb{R}^m \rightarrow \mathbb{R}$  convex, Lipschitz and bounded from below, it holds*

$$(6.10) \quad \|e^{Q_t \varphi}\|_{a+\rho' t} \leq \|e^\varphi\|_a.$$

Moreover

- (i)  $\Rightarrow$  (ii) for all  $\lambda \in (0, 1/b)$  and with  $C = 1/(1 - b\lambda)$ ;
- (ii)  $\Rightarrow$  (iii) with  $\rho = \frac{\lambda}{C}$ ;
- (iii)  $\Rightarrow$  (iv) with  $\rho' = \rho$ ;
- (iv)  $\Rightarrow$  (i) with  $b = \frac{1}{\rho'}$ .

**Remark 6.11.** *The implication (ii)  $\Rightarrow$  (i) is a variant of a well known result due to Otto and Villani [26] showing that the logarithmic Sobolev inequality implies the classical transport-entropy inequality  $\mathbf{T}_2$ . Here we will make use of the arguments developed in [5]. On the other hand, in the classical setting, the equivalence (i)  $\iff$  (ii) was studied and developed in [12, 13, 14].*

Observe that the relations between the various constants is almost optimal. Indeed, starting from  $\overline{\mathbf{T}}_2^-(b)$ , we deduce from (ii)  $\Rightarrow$  (iii) that the log-Sobolev inequality (6.9) holds with  $\rho = \sup_{\lambda \in (0, 1/b)} \lambda/C = \sup_{\lambda \in (0, 1/b)} \lambda(1 - b\lambda) = \frac{1}{4b}$  (the maximum is reached at  $\lambda = 1/(2b)$ ). From this we deduce (iv) with  $\rho' = 1/(4b)$  which gives back  $\overline{\mathbf{T}}_2^-(4b)$ , and in all we are off only by a factor 4.

We may make use of the above result to obtain example of measures satisfying  $\overline{\mathbf{T}}_2^-(b)$  in Section 7. Indeed, the ‘‘convex’’ log-Sobolev inequality (6.9) was studied in the literature [1].

We will use the following classical smoothing property of the infimum convolution operator.

**Lemma 6.12.** *Let  $\|\cdot\|$  be a norm on  $\mathbb{R}^m$  whose dual norm is strictly convex. If  $\varphi : \mathbb{R}^m \rightarrow \mathbb{R}$  is a convex function, then for all  $t > 0$ , the function  $Q_t\varphi$  defined by*

$$Q_t\varphi(x) = \inf_{y \in \mathbb{R}^m} \left\{ \varphi(y) + \frac{1}{2t} \|x - y\|^2 \right\}, \quad x \in \mathbb{R}^m.$$

*is also convex and  $\mathcal{C}^1$ -smooth on  $\mathbb{R}^m$ .*

*Proof.* The fact that  $Q_t\varphi$  is convex is well known and easy to check. Consider the Fenchel-Legendre transform of  $Q_t\varphi$  defined by  $(Q_t\varphi)^*(x) = \sup_{y \in \mathbb{R}^m} \{x \cdot y - Q_t\varphi(y)\}$ . A simple calculation shows that  $(Q_t\varphi)^*(x) = \varphi^*(x) + \frac{1}{2} \|x\|_*^2$ , for all  $x \in \mathbb{R}^m$ . By assumption,  $\|\cdot\|_*$  satisfies (6.6). This easily implies (and is actually equivalent to) that the convex function  $x \mapsto \|x\|_*^2$  is strictly convex (in the usual sense : if  $x \neq y$ , then  $\|(1-t)x + ty\|_*^2 < (1-t)\|x\|_*^2 + t\|y\|_*^2$ , for all  $t \in (0, 1)$ ). Therefore, the function  $x \mapsto (Q_t\varphi)^*(x)$  is strictly convex on  $\mathbb{R}^m$ . A classical result in Fenchel-Legendre duality (see e.g. [17, Theorem E.4.1.1]) then implies that  $(Q_t\varphi)^{**} = Q_t\varphi$  is  $\mathcal{C}^1$ -smooth on  $\mathbb{R}^m$ .  $\square$

*Proof of Theorem 6.8.* That (i) implies (ii) is given in Proposition 6.3.

To prove that (ii) implies (iii), fix  $\varphi : \mathbb{R}^m \rightarrow \mathbb{R}$  a  $\mathcal{C}^1$ -smooth function which is convex, Lipschitz and bounded from below. Then, by convexity, for all  $x, y \in \mathbb{R}^m$ , it holds

$$\varphi(x) - \varphi(y) \leq \nabla\varphi(x) \cdot (x - y).$$



where  $u \cdot v$  denotes the scalar product of  $u, v \in \mathbb{R}^m$ . Hence, given  $\lambda > 0$  and  $x \in \mathbb{R}^m$ , by the Cauchy-Schwarz inequality  $u \cdot v \leq \frac{1}{2\lambda} \|u\|_*^2 + \frac{\lambda}{2} \|v\|^2$ ,  $u, v \in \mathbb{R}^m$ , we have

$$\begin{aligned} \varphi(x) - R_c^\lambda \varphi(x) &= \sup_{p \in \mathcal{P}_1(\mathbb{R}^m)} \left\{ \int [\varphi(x) - \varphi(y)] p(dy) - \frac{\lambda}{2} \|x - \int y p(dy)\|^2 \right\} \\ &\leq \sup_{p \in \mathcal{P}_1(\mathbb{R}^m)} \left\{ \int \nabla \varphi(x) \cdot (x - y) p(dy) - \frac{\lambda}{2} \|x - \int y p(dy)\|^2 \right\} \\ &= \sup_{p \in \mathcal{P}_1(\mathbb{R}^m)} \left\{ \nabla \varphi(x) \cdot (x - \int y p(dy)) - \frac{\lambda}{2} \|x - \int y p(dy)\|^2 \right\} \\ &\leq \frac{1}{2\lambda} \|\nabla \varphi(x)\|_*^2. \end{aligned}$$

The expected result follows.

To prove that (iii) implies (iv), we follow the now classical argument from [5] based on the Hamilton-Jacobi equation satisfied by  $(t, x) \mapsto Q_t \varphi(x)$ . Since we do not assume that  $\mu$  is absolutely continuous with respect to Lebesgue measure (one of our main motivations is to study transport inequalities for *discrete* measures), there are some technical difficulties to clarify in order to adapt the proof of [5, Theorem 2.1] to our framework. First, as shown in [14] or [3], the following Hamilton-Jacobi equation holds for *all*  $t > 0$  and  $x \in \mathbb{R}^m$  :

$$(6.13) \quad \frac{d^+}{dt} Q_t \varphi(x) = -\frac{1}{2} |\nabla^- Q_t \varphi|^2(x),$$

where,  $d^+/dt$  stands for the right derivative, and by definition  $|\nabla^- f|(x)$  is a notation for the *local slope* of a function  $f$  at a point  $x$ , defined by

$$|\nabla^- f|(x) = \limsup_{y \rightarrow x} \frac{[f(y) - f(x)]_-}{\|y - x\|}.$$

Here, since  $\varphi$  is *convex*, the regularization property of the inf-convolution operator  $Q_t$  given in Lemma 6.12 implies that for all  $t > 0$ , the function  $x \mapsto Q_t \varphi(x)$  is actually  $\mathcal{C}^1$ -smooth on  $\mathbb{R}^m$ . It is then easily checked that  $|\nabla^- Q_t \varphi|(x) = \|\nabla Q_t \varphi(x)\|_*$ . Moreover, according to Lemma 6.12 again, if  $\varphi: \mathbb{R}^m \rightarrow \mathbb{R}$  is convex, then so does  $Q_t \varphi$ . Therefore, (6.9) can be applied to the function  $Q_t \varphi$  for all  $t > 0$ . To complete the proof of the implication, we leave it to the reader to follow the proof of [5, Theorem 2.1] (see also [14, Theorem 1.11]).

Finally we prove that (iv) implies (i). We observe that, at  $t = 1$  and  $a = 0$ , (6.10) precisely means that,

$$\int e^{\rho' Q_1 \varphi} d\mu \leq e^{\rho' \int \varphi d\mu}.$$

This is equivalent to  $\overline{\mathbf{T}}_2^-(1/\rho')$ , thanks to Proposition 5.5 and to the fact that, as recalled above,  $Q_1 \varphi = R_c \varphi = \inf_{p \in \mathcal{P}_1(X)} \{ \int \varphi(y) p(dy) + c(x, p) \}$ , for any  $\varphi: \mathbb{R}^m \rightarrow \mathbb{R}$  convex, Lipschitz and bounded from below. This completes the proof.  $\square$

In order to give a series of equivalent formulations of  $\overline{\mathbf{T}}_2^+(b)$ , we need to introduce the notion of  $c$ -convexity (see e.g. [37]). We recall that if  $c: X \times X$  is some cost function on a space  $X$ , a function  $f: X \rightarrow \mathbb{R} \cup \{\pm\infty\}$  is said  $c$ -convex if there exists some function  $g: X \rightarrow \mathbb{R} \cup \{\pm\infty\}$  such that

$$f(x) = \sup_{y \in X} \{g(y) - c(x, y)\}, \quad \forall x \in X.$$

In what follows, we will use this notion with  $c(x, y) = \frac{\lambda}{2} \|x - y\|^2$ ,  $x, y \in \mathbb{R}^m$ , where  $\lambda > 0$  and  $\|\cdot\|$  is some norm on  $X = \mathbb{R}^m$ , such that  $\|\cdot\|_*$  is a strictly convex norm in the sense of (6.6). In other words, a function  $f: \mathbb{R}^m \rightarrow \mathbb{R} \cup \{\pm\infty\}$  is  $\frac{\lambda}{2} \|\cdot\|^2$ -convex if there exists  $g: \mathbb{R}^m \rightarrow \mathbb{R} \cup \{\pm\infty\}$  such that  $f = P_{\frac{1}{\lambda}} g$  (recall the definition of  $P_t$  from 6.5). In

e.g. [14, Proposition 2.2] it is proved that  $f$  is  $\frac{\lambda}{2}\|\cdot\|^2$ -convex if and only if  $f = P_{\frac{1}{\lambda}}Q_{\frac{1}{\lambda}}f$ . Furthermore (see e.g. [14, Proposition 2.3]), if  $f$  is of class  $\mathcal{C}^2$  and  $\|\cdot\| = |\cdot|$  is the Euclidean norm, then  $f$  is  $\frac{\lambda}{2}|\cdot|^2$ -convex if and only if  $\text{Hess}f \geq -\lambda\text{Id}$  (as a matrix), where  $\text{Hess}$  denotes the Hessian.

To avoid the use of too heavy a terminology, we will denote by  $\mathcal{F}_\lambda(\mathbb{R}^m)$ ,  $\lambda > 0$ , the class of all functions  $f : \mathbb{R}^m \rightarrow \mathbb{R}$  that are *concave*, *Lipschitz*, *bounded from above* and  $\frac{\lambda}{2}\|\cdot\|^2$ -convex.

**Remark 6.14.** According to Lemma 6.12, if  $g$  is concave on  $\mathbb{R}^m$  and  $\lambda > 0$ , then  $Q_{1/\lambda}(-g)$  is convex and  $\mathcal{C}^1$ -smooth. In particular,  $f = -Q_{1/\lambda}(-g)$  is concave and  $\mathcal{C}^1$ -smooth. But  $f = -Q_{1/\lambda}(-g) = P_{1/\lambda}(g)$  and thus  $f$  is also  $\frac{\lambda}{2}\|\cdot\|^2$ -convex. Furthermore, if  $g$  is assumed to be Lipschitz and bounded from above, then  $f$  is also Lipschitz and bounded from above. This shows that the class  $\mathcal{F}_\lambda(\mathbb{R}^m) \cap \mathcal{C}^1(\mathbb{R}^m)$  is not empty.

**Theorem 6.15.** Suppose that  $\|\cdot\|_*$  is a strictly convex norm and let  $\mu \in \mathcal{P}_1(\mathbb{R}^m)$ . Then the following are equivalent:

- (i) there exists  $b > 0$  such that  $\overline{\mathbf{T}}_2^+(b)$  holds;
- (ii) there exist  $\lambda, C > 0$  such that for all  $\varphi \in \mathcal{F}_\lambda(\mathbb{R}^m)$ , it holds

$$(6.16) \quad \text{Ent}_\mu(e^\varphi) \leq C \int (\varphi - Q_{1/\lambda}\varphi)e^\varphi d\mu;$$

- (iii) there exist  $\rho, \lambda' > 0$  such that for all  $\mathcal{C}^1$ -smooth function  $\varphi \in \mathcal{F}_{\lambda'}(\mathbb{R}^m)$ , it holds

$$(6.17) \quad \text{Ent}_\mu(e^\varphi) \leq \frac{1}{2\rho} \int \|\nabla\varphi\|_*^2 e^\varphi d\mu.$$

Moreover

- (i)  $\Rightarrow$  (ii) for all  $\lambda \in (0, 1/b)$  and with  $C = 1/(1 - b\lambda)$ ;
- (ii)  $\Rightarrow$  (iii) for all  $\lambda' \in (0, \lambda)$  and with  $\rho = \frac{\lambda - \lambda'}{C}$ ;
- (iii)  $\Rightarrow$  (i) with  $b = \frac{\rho + \lambda'}{\rho\lambda'}$ .

**Remark 6.18.** Also, Equation (6.16) is very close to (yet different from) the  $(\tau)$ -log-Sobolev inequality (6.2). The difference is coming from the fact that, for concave functions,  $R_c f \neq Qf$ , while equality holds for convex functions.

In particular, we emphasise the fact that  $\overline{\mathbf{T}}_2^-(b)$  encompasses information about convex functions, while  $\overline{\mathbf{T}}_2^+(b)$  about concave functions.

Finally, we observe that the constants in the various implications are almost optimal. Indeed, starting from  $\overline{\mathbf{T}}_2^+(b)$ , we end up with  $\overline{\mathbf{T}}_2^+(b')$ , with  $b' = \frac{(\lambda - \lambda')(1 - b\lambda) + \lambda'}{\lambda'(\lambda - \lambda')(1 - b\lambda)}$  with  $\lambda \in (0, 1/b)$  and  $\lambda' \in (0, \lambda)$ . Choosing  $\lambda = 1/(2b)$  and  $\lambda' = 1/(4b)$  one gets  $b' = 12b$  and we are off only by a factor at most 12.

*Proof.* To prove that (i) implies (ii), we follow the argument of the proof of Proposition 6.3. Consider a concave function  $f$ , Lipschitz and bounded above,  $\lambda \in (0, 1/b)$  and define for simplicity  $s = 1/\lambda$  and  $d\nu_f = \frac{\exp\{P_s f\}}{\int \exp\{P_s f\} d\mu} d\mu$ . By Jensen's Inequality we have

$$\begin{aligned} H(\nu_f|\mu) &= \int \log \left( \frac{e^{P_s f}}{\int e^{P_s f} d\mu} \right) \frac{e^{P_s f}}{\int e^{P_s f} d\mu} d\mu = \int P_s f d\nu_f - \log \int e^{P_s f} d\mu \\ &\leq \int P_s f d\nu_f - \int P_s f d\mu = \int [P_s f - f] d\nu_f - \int P_s f d\mu + \int f d\nu_f \\ &\leq \int [P_s f - f] d\nu_f + \lambda \overline{\mathbf{T}}_2(\nu_f|\mu). \end{aligned}$$

where in the last line we used the homogeneity of the transport cost (as a function of the cost (recall that  $s = 1/\lambda$ )) and the duality theorem (Corollary 3.11) to ensure that (since  $Q_1(-\varphi) = -P_1\varphi$ )

$$\begin{aligned}\bar{\mathcal{T}}_2(\nu_f|\mu) &= \sup \left\{ \int Q_1\varphi d\mu - \int \varphi d\nu_f; \varphi \text{ convex, Lipschitz, bounded from below} \right\} \\ &= \sup \left\{ - \int P_1\varphi d\mu + \int \varphi d\nu_f; \varphi \text{ concave, Lipschitz, bounded from above} \right\}.\end{aligned}$$

Applying  $\bar{\mathcal{T}}_2^+(b)$  and rearranging the terms, we end up with the following inequality (since  $(\int \exp\{P_s f\} d\mu) H(\nu_f|\mu) = \text{Ent}_\mu(\exp\{P_s f\})$ ):

$$\text{Ent}_\mu(e^{P_s f}) \leq \frac{1}{1 - \lambda b} \int [P_s f - f] e^{P_s f} d\mu,$$

which holds for any  $f$  concave, Lipschitz and bounded above, and for any  $\lambda \in (0, 1/b)$  and  $s = 1/\lambda$ . Now, our aim is to get rid of  $P_s f$ . To that purpose, we observe that, since  $f$  is concave, Lipschitz and bounded above,  $Q_s f$  is also concave, Lipschitz and bounded above<sup>4</sup> (for any  $s \geq 0$ ), so that, if we assume in addition that  $f$  is  $\frac{\lambda}{2}\|\cdot\|^2$ -convex, applying the latter to  $Q_s f$  and using that  $P_s Q_s f = f$ , we finally get the desired result of Item (ii).

Now we prove that (ii) implies (iii). Assume Item (ii) and consider a function  $f \in \mathcal{F}_{\lambda'}(\mathbb{R}^m)$ , with  $\lambda' \in (0, \lambda)$ . Our aim is to make use of the  $\frac{\lambda'}{2}\|\cdot\|^2$ -convexity property of  $f$  to bound  $f - Q_{1/\lambda} f$  from above by  $\|\nabla f\|_*^2$ ; we may follow [14].

Since  $f$  is  $\frac{\lambda'}{2}\|\cdot\|^2$ -convex, it satisfies  $P_s Q_s f = f$ , where for simplicity  $s = 1/\lambda'$  (see e.g. [14, Proposition 2.2]). Define  $m(x) = \{\bar{y} \in \mathbb{R}^m : f(x) = g(\bar{y}) - \frac{\lambda'}{2}\|x - \bar{y}\|^2\}$  (i.e. the set of points where the supremum is reached (that is non empty by simple compactness arguments (see [14, Lemma 2.6])). Given  $\bar{y} \in m(x)$ , we have for all  $z \in \mathbb{R}^m$ ,

$$(6.19) \quad f(x) = Q_s f(\bar{y}) - \frac{\lambda'}{2}\|x - \bar{y}\|^2 \leq f(z) + \frac{\lambda'}{2}(\|z - \bar{y}\|^2 - \|x - \bar{y}\|^2).$$

Since  $f$  is concave and  $\mathcal{C}^1$ -smooth, it holds

$$f(z) \leq f(x) + \nabla f(x) \cdot (z - x), \quad \forall z \in \mathbb{R}^m.$$

Inserting this inequality in (6.19), one gets

$$0 \leq \nabla f(x) \cdot (z - x) + \frac{\lambda'}{2}(\|z - \bar{y}\|^2 - \|x - \bar{y}\|^2), \quad \forall z \in \mathbb{R}^m.$$

Applying this to  $z_t = (1 - t)x + t\bar{y}$ , with  $t \in (0, 1)$ , one obtains

$$0 \leq t \nabla f(x) \cdot (\bar{y} - x) + \frac{\lambda'}{2}((1 - t)^2 - 1)\|x - \bar{y}\|^2.$$

Dividing by  $t$  and letting  $t \rightarrow 0$ , one ends up with the inequality

$$\lambda' \|x - \bar{y}\|^2 \leq \nabla f(x) \cdot (\bar{y} - x) \leq \|\nabla f(x)\|_* \|x - \bar{y}\|.$$

<sup>4</sup>These facts follow from the fact that  $Q_s f(x) = \inf_y \{f(x - y) + \frac{s}{2}\|y\|^2\}$ . Hence  $Q_s f$  is concave as infimum of concave functions. On the other hand,  $x \mapsto f(x - y) + \frac{s}{2}\|y\|^2$  are uniformly (in  $y$ ) Lipschitz functions so that  $Q_s f$  is again Lipschitz as infimum of Lipschitz functions. Finally,  $Q_s f \leq f$  and therefore is bounded above.

According to (6.19), the triangle inequality, and the inequality  $\|x - \bar{y}\| \leq \frac{1}{\lambda'} \|\nabla f(x)\|_*$ , one gets

$$\begin{aligned} f(x) &\leq f(z) + \frac{\lambda'}{2} \left( \|z - x\|^2 + 2\|z - x\| \|x - \bar{y}\| \right) \\ &\leq f(z) + \frac{\lambda'}{2} \left( \|z - x\|^2 + 2\|z - x\| \frac{\|\nabla f(x)\|_*}{\lambda'} \right) \\ &\leq f(z) + \frac{\lambda}{2} \|z - x\|^2 + \left( \|z - x\| \|\nabla f(x)\|_* - \frac{\lambda - \lambda'}{2} \|z - x\|^2 \right) \\ &\leq f(z) + \frac{\lambda}{2} \|z - x\|^2 + \frac{1}{2(\lambda - \lambda')} \|\nabla f(x)\|_*^2. \end{aligned}$$

Optimizing over  $z \in \mathbb{R}^m$ , one gets the inequality

$$f(x) - Q_{1/\lambda} f(x) \leq \frac{1}{2(\lambda - \lambda')} \|\nabla f(x)\|_*^2,$$

which inserted into (6.16) yields to (6.17).

It remains to prove that (iii) implies (i). To that purpose, let  $\ell(t) := -\rho(1-t)$ ,  $t \in (0, 1)$  (observe that  $\ell(t) \leq 0$ ), set  $s = -\ell(t)/\lambda'$ , and consider a convex, Lipschitz and bounded below function  $f: \mathbb{R}^m \rightarrow \mathbb{R}$ . We shall apply the log-Sobolev inequality to  $\varphi = \ell(t)Q_t f$  for a given  $t \in (0, 1)$ . We need first to verify that  $\varphi$  is concave, Lipschitz, bounded above and  $\lambda'c$ -convex. Since  $f$  is convex,  $Q_t f$  is convex and so, since  $\ell(t) \leq 0$ ,  $\varphi$  is concave. On the other hand, since  $f$  is Lipschitz, so does  $\varphi$ . Also,  $f$  being bounded below,  $Q_t f \geq \inf f$  and  $\ell(t) \leq 0$ , we have  $\varphi = \ell(t)Q_t f \leq \ell(t) \inf f$  which proves that  $\varphi$  is bounded above. Finally, since  $Q_t$  is a semi-group and since in general  $Q_u(g) = -P_u(-g)$ , we have for all  $t \in (\frac{\rho}{\rho+\lambda'}, 1)$  (to ensure that  $s \leq t$ ),

$$\varphi = \ell(t)Q_s(Q_{t-s}f) = -\ell(t)P_s(-Q_{t-s}f) = P_{-\frac{s}{\ell(t)}}(\ell(t)Q_{t-s}f) = P_{\frac{1}{\lambda'}}(\ell(t)Q_{t-s}f),$$

hence  $\varphi$  is  $\lambda'c$ -convex. In turn, applying the log-Sobolev inequality to  $\varphi$  (which is  $\mathcal{C}^1$ -smooth according to Lemma 6.12), we end up with the following inequality that we shall use later on:

$$\int \ell(t)Q_t f e^{\ell(t)Q_t f} d\mu - H(t) \log H(t) = \text{Ent}_\mu(e^{\ell(t)Q_t f}) \leq \frac{\ell(t)^2}{2\rho} \int \|\nabla Q_t f\|_*^2 e^{\ell(t)Q_t f} d\mu,$$

where  $H(t) := \int e^{\ell(t)Q_t f} d\mu$ . Hence, by the Hamilton-Jacobi equation (6.13),

$$\begin{aligned} \frac{d^+}{dt} \left( \frac{1}{\ell(t)} \log H(t) \right) &= \frac{1}{\ell(t)^2 H(t)} \left( -\ell'(t) H(t) \log H(t) + \ell(t) H'(t) \right) \\ &= \frac{1}{\ell(t)^2 H(t)} \left( \ell'(t) \text{Ent}_\mu(e^{\ell(t)Q_t f}) + \ell(t)^2 \int \frac{\partial Q_t f}{\partial t} e^{\ell(t)Q_t f} d\mu \right) \\ &= \frac{\ell'(t)}{\ell(t)^2 H(t)} \left( \text{Ent}_\mu(e^{\ell(t)Q_t f}) + \frac{\ell(t)^2}{2\ell'(t)} \int \|\nabla Q_t f\|_*^2 e^{\ell(t)Q_t f} d\mu \right) \\ &\leq \frac{\ell'(t)}{2H(t)} \left( \frac{1}{\rho} - \frac{1}{\ell'(t)} \right) \int \|\nabla Q_t f\|_*^2 e^{\ell(t)Q_t f} d\mu = 0 \end{aligned}$$

since  $\ell'(t) = \rho$ . Therefore the function  $t \mapsto \|e^{Q_t f}\|_{\ell(t)}$  is non-increasing on  $(\frac{\rho}{\rho+\lambda'}, 1)$ . In

particular, in the limit, we get  $\|e^{Q_1 f}\|_{\ell(1)} \leq \left\| e^{Q_{\frac{\rho}{\rho+\lambda'}} f} \right\|_{\ell(\frac{\rho}{\rho+\lambda'})}$  that we can rephrase as

$$e \int Q_1 f d\mu \left( \int e^{-\frac{\rho\lambda'}{\rho+\lambda'}} Q_{\frac{\rho}{\rho+\lambda'}} f d\mu \right)^{\frac{\rho+\lambda'}{\rho\lambda'}} \leq 1.$$

Now, since  $Q_u f \leq f$ , we conclude that

$$e^{\int Q_1 f d\mu} \left( \int e^{-\frac{\rho\lambda'}{\rho+\lambda'} f} d\mu \right)^{\frac{\rho+\lambda'}{\rho\lambda'}} \leq 1,$$

which implies  $\overline{\mathbf{T}}_2^+(\frac{\rho+\lambda'}{\rho\lambda'})$  by Proposition 5.5 and Corollary 3.11. This completes the proof.  $\square$

## 7. EXAMPLES

In this section, we give some examples of probability measures satisfying weak transport inequalities. We start with the Bernoulli measure, from which we derive weak transport inequalities for the binomial law and the Poisson distribution. Also, using the characterization of  $\overline{\mathbf{T}}_2^-$  by means of log-Sobolev type inequalities of the previous section and results from [1], we may give more examples of measures satisfying such a transport-entropy inequality on the line.

We will use some results for the Bernoulli measure, derived in [30], and as such introduce some notations from there.

Let  $w: \mathbb{R} \rightarrow \mathbb{R}_+ \cup \{+\infty\}$  be defined as

$$w(x) = \begin{cases} (1-x) \log(1-x) + x & \text{if } x \leq 1 \\ +\infty & \text{if } x > 1, \end{cases}$$

and observe that  $(1-x) \log(1-x) + x$  compares to  $\min(x^2, |x| \log(1+|x|))$ . Then, given  $\rho \in (0, 1)$ , define  $u_\rho: \mathbb{R} \rightarrow \mathbb{R}_+ \cup \{+\infty\}$  as

$$\begin{aligned} u_\rho(x) &= \frac{1-\rho}{\rho} w\left(-\frac{\rho x}{1-\rho}\right) + w(x) \\ &= \begin{cases} \frac{1-\rho(1-x)}{\rho} \log \frac{1-\rho(1-x)}{1-\rho} + (1-x) \log(1-x), & \text{if } -\frac{1-\rho}{\rho} \leq x \leq 1 \\ +\infty & \text{otherwise} \end{cases} \end{aligned}$$

and given  $t \in (0, 1)$ , let  $\theta_{\rho,t}: \mathbb{R} \rightarrow \mathbb{R}$  be defined as

$$\theta_{\rho,t}(h) = \frac{1}{t(1-t)} \inf_{\tau \geq 1} \begin{cases} \frac{1}{\tau} [(1-t)u_\rho(1-\tau) + tu_\rho(1-\tau(1-h))] & \text{if } h \geq 0 \\ \frac{1}{\tau} [(1-t)u_{1-\rho}(1-\tau) + tu_{1-\rho}(1-\tau(1+h))] & \text{if } h < 0. \end{cases}$$

It is easy to see that  $w$  and  $u_\rho$  are convex. We shall prove in Appendix A that  $\theta_{\rho,t}$  is also convex for any  $\rho$  and  $t$  and compares to  $h^2$  on  $[-1, 1]$  and  $\theta_{\rho,t} \equiv +\infty$  on  $\mathbb{R} \setminus [-1, 1]$  (see [30, Proposition 2.4] and Appendix A). Finally, we define  $\theta_{\rho,0}$  and  $\theta_{\rho,1}$  as the limit when  $t$  tends to 0, respectively 1, of  $\theta_{\rho,t}$ . As proved in [30, Proposition 2.4], it holds

$$\theta_{\rho,0}(h) = \begin{cases} u_\rho(h) & \text{if } h \geq 0 \\ u_{1-\rho}(-h) & \text{if } h < 0, \end{cases}$$

and

$$\theta_{\rho,1}(h) = \begin{cases} \frac{1}{\rho} \left[ (1-\rho-h) \log \frac{1-\rho-h}{1-\rho} - (1-h) \log(1-h) \right] & \text{if } h \in [0, 1-\rho] \\ \frac{1}{1-\rho} \left[ (\rho+h) \log \frac{\rho+h}{\rho} - (1+h) \log(1+h) \right] & \text{if } h \in [-\rho, 0] \\ +\infty & \text{otherwise.} \end{cases}$$

Finally we define  $\tilde{\theta}_{\rho,t}(h): \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$  by  $\tilde{\theta}_{\rho,t}(h) := \min(\theta_{\rho,t}(h), \theta_{\rho,t}(-h))$ . It is a tedious but easy exercise to verify that, for all  $t \in [0, 1]$  and all  $h \in \mathbb{R}$ ,  $\tilde{\theta}_{\rho,t}(h) = \theta_{\rho,t}(|h|)$  when  $\rho \in (0, 1/2]$  and  $\tilde{\theta}_{\rho,t}(h) = \theta_{1-\rho,t}(|h|)$  when  $\rho \in [1/2, 1)$ .

**Warning:** In this section, for simplicity and when there is no confusion, we shall drop the index  $r = 0$  in many notations, as for example  $\mathcal{P}(\cdot) = \mathcal{P}_0(\cdot)$  and  $\Phi(\cdot) = \Phi_0(\cdot)$ . This is justified by the fact that, on any finite graph,  $\mathcal{P}_0(\cdot)$  coincides with the set of all probability measures  $\mathcal{P}(\cdot)$ , and similarly  $\Phi_0(\cdot)$  coincides with the set of all functions.

**7.1. Weak transport inequality for the Bernoulli measure and the product of Bernoulli measures.** Using a result from [30] and the duality results proved in Section 3, we get the following weak transport inequalities for the non symmetric Bernoulli measure. Set  $\mu_\rho := (1 - \rho)\delta_0 + \rho\delta_1$ ,  $\rho \in [0, 1]$ .

**Proposition 7.1.** *For all  $t, \rho \in (0, 1)$ , it holds*

$$(7.2) \quad \overline{\mathcal{T}}_{\theta_{\rho,t}}(\nu_1|\nu_2) \leq \frac{1}{1-t}H(\nu_1|\mu_\rho) + \frac{1}{t}H(\nu_2|\mu_\rho) \quad \forall \nu_1, \nu_2 \in \mathcal{P}(\{0, 1\}).$$

In particular

$$\overline{\mathcal{T}}_{\theta_{\rho,1}}(\mu_\rho|\nu) \leq H(\nu|\mu_\rho) \quad \text{and} \quad \overline{\mathcal{T}}_{\theta_{\rho,0}}(\nu|\mu_\rho) \leq H(\nu|\mu_\rho) \quad \forall \nu \in \mathcal{P}(\{0, 1\}).$$

Moreover, the same inequalities hold replacing  $\overline{\mathcal{T}}_{\theta_{\rho,t}}$ ,  $\overline{\mathcal{T}}_{\theta_{\rho,1}}$  and  $\overline{\mathcal{T}}_{\theta_{\rho,0}}$  respectively by  $\widetilde{\mathcal{T}}_{\tilde{\theta}_{\rho,t}}$ ,  $\widetilde{\mathcal{T}}_{\tilde{\theta}_{\rho,1}}$  and  $\widetilde{\mathcal{T}}_{\tilde{\theta}_{\rho,0}}$ .

The proof of the above proposition relies on the following lemma<sup>5</sup>.

**Lemma 7.3** ([30]). *For all  $t, \rho \in (0, 1)$  and all convex functions  $f: \mathbb{R} \rightarrow \mathbb{R}$ , it holds*

$$\left( \int e^{tQ_{\theta_{\rho,t}}f} d\mu_\rho \right)^{1/t} \left( \int e^{-(1-t)f} d\mu_\rho \right)^{1/(1-t)} \leq 1,$$

where

$$Q_{\theta_{\rho,t}}f(x) := \inf_{y \in \mathbb{R}} \{f(y) + \theta_{\rho,t}(x - y)\}.$$

*Proof of Proposition 7.1.* Inequality (7.2) follows from Lemma 7.3 and the Bobkov-Götze dual characterization (Proposition 5.5) together with Corollary 3.11. The two other inequalities follow by choosing  $\nu_1 = \mu_\rho$  and  $\nu_2 = \mu_\rho$ , and by taking the limit  $t \rightarrow 1$  and  $t \rightarrow 0$ , respectively.

Now, observe that for every probability measure  $p \in \mathcal{P}(\{0, 1\})$  and for any  $x \in \{0, 1\}$ , it holds

$$\int |x - y| p(dy) = p(\{1 - x\}) = \left| x - \int y p(dy) \right|.$$

Hence, since  $\tilde{\theta}_{\rho,t}$  is increasing on  $[0, +\infty)$ ,  $\widetilde{\mathcal{T}}_{\tilde{\theta}_{\rho,t}}(\nu_1|\nu_2) \leq \overline{\mathcal{T}}_{\theta_{\rho,t}}(\nu_1|\nu_2)$  and the results involving  $\widetilde{\mathcal{T}}_{\tilde{\theta}_{\rho,t}}$ ,  $\widetilde{\mathcal{T}}_{\tilde{\theta}_{\rho,1}}$  and  $\widetilde{\mathcal{T}}_{\tilde{\theta}_{\rho,0}}$  follow. The proof is complete.  $\square$

By Theorem 5.11, the weak transport inequalities for the Bernoulli measure  $\mu_\rho$  given in Proposition 7.1 tensorises. Hence, the product of Bernoulli measures  $\mu_\rho^n := \mu_\rho \otimes \cdots \otimes \mu_\rho$  on the hypercube  $\Omega_n = \{0, 1\}^n$  satisfies the following  $n$ -dimensional version of the  $\overline{\mathcal{T}}$  and  $\widetilde{\mathcal{T}}$ -transport-entropy inequalities. Recall that the corresponding  $n$ -dimensional costs are defined, for all  $x = (x_1, \dots, x_n) \in \Omega_n$  and all  $p \in \mathcal{P}(\Omega_n)$ , respectively by

$$\bar{c}_{\rho,t}^{(n)}(x, q) := \sum_{i=1}^n \theta_{\rho,t} \left( x_i - \int_{\Omega_1} y_i q_i(dy_i) \right)$$

<sup>5</sup>To be precise the result of Lemma 7.3 is proved for the centered non symmetric Bernoulli measure  $(1 - \rho)\delta_{-\rho} + \rho\delta_{1-\rho}$ ,  $\rho \in (0, 1)$ . However, by a simple translation argument the result also holds for the Bernoulli measure  $\mu_\rho$  (details are left to the reader).

and

$$\tilde{c}_{\rho,t}^{(n)}(x, q) := \sum_{i=1}^n \tilde{\theta}_{\rho,t} \left( \int_{\Omega_1} |x_i - y_i| q_i(dy_i) \right),$$

where  $q_i \in \mathcal{P}(\Omega_1)$  is the  $i$ -th marginal of  $q$ . We denote by  $\overline{\mathcal{T}}_{\tilde{c}_{\rho,t}^{(n)}}$  and  $\widetilde{\mathcal{T}}_{\tilde{c}_{\rho,t}^{(n)}}$  the corresponding transport costs. Applying Theorem 5.11, we immediately get, from Proposition 7.1, the following weak transport-entropy inequalities for product of Bernoulli measures.

**Corollary 7.4.** *For all  $t \in (0, 1)$ , all  $\rho \in (0, 1)$  and all  $n = 1, 2, \dots$ , it holds*

$$(7.5) \quad \overline{\mathcal{T}}_{\tilde{c}_{\rho,t}^{(n)}}(\nu_1 | \nu_2) \leq \frac{1}{1-t} H(\nu_1 | \mu_\rho^n) + \frac{1}{t} H(\nu_2 | \mu_\rho^n) \quad \forall \nu_1, \nu_2 \in \mathcal{P}(\Omega_n).$$

*In particular*

$$\overline{\mathcal{T}}_{\tilde{c}_{\rho,1}^{(n)}}(\mu_\rho^n | \nu) \leq H(\nu | \mu_\rho^n) \quad \text{and} \quad \overline{\mathcal{T}}_{\tilde{c}_{\rho,0}^{(n)}}(\nu | \mu_\rho^n) \leq H(\nu | \mu_\rho^n) \quad \forall \nu \in \mathcal{P}(\Omega_n).$$

*Moreover, the same inequalities hold replacing  $\overline{\mathcal{T}}_{\tilde{c}_{\rho,t}^{(n)}}$ ,  $\overline{\mathcal{T}}_{\tilde{c}_{\rho,1}^{(n)}}$  and  $\overline{\mathcal{T}}_{\tilde{c}_{\rho,0}^{(n)}}$  respectively by  $\widetilde{\mathcal{T}}_{\tilde{c}_{\rho,t}^{(n)}}$ ,  $\widetilde{\mathcal{T}}_{\tilde{c}_{\rho,1}^{(n)}}$  and  $\widetilde{\mathcal{T}}_{\tilde{c}_{\rho,0}^{(n)}}$ .*

**7.2. Weak transport cost for the binomial law.** In this section we prove weak transport cost inequalities for the binomial distribution  $B(n, \rho)$ ,  $\rho \in (0, 1)$ . The basic idea is to project the  $n$ -dimensional transport cost inequalities (7.5), from the hypercube  $\Omega_n = \{0, 1\}^n$  onto  $I_n := \{0, 1, \dots, n\}$ , the state space of  $B(n, \rho)$ .

Let  $\mu_{n,\rho}$  denote the binomial measure on  $I_n$ , i.e.  $\mu_{n,\rho}(k) = \binom{n}{k} \rho^k (1-\rho)^{n-k}$  for all  $k \in I_n$ . Then the image measure of  $\mu_\rho^n$  by the projection  $\varphi: \Omega_n \ni (x_1, \dots, x_n) \mapsto \sum_{i=1}^n x_i \in I_n$  is the measure  $\mu_{n,\rho}$ . Let also  $\Omega_n^k := \{x \in \Omega_n: \phi(x) = k\}$  be the slices of the cube,  $k \in I_n$ .

We may start with a general projection result. The lemma below shows that any  $n$ -dimensional weak cost of type  $c^n$  on the hypercube provides a weak cost on  $I_n$  through the projection  $\varphi$ . As a consequence, any weak transport-entropy inequality on the hypercube give rise to a weak transport-entropy in  $I_n$  with the projected cost.

**Lemma 7.6.** *Let  $c: \Omega_1 \times \mathcal{P}(\Omega_1) \rightarrow [0, \infty]$  be a cost function and construct  $c^{(n)}: \Omega_n \times \mathcal{P}(\Omega_n) \rightarrow [0, \infty]$  on  $\Omega_n$  by*

$$c^{(n)}(x, q) := \sum_{i=1}^n c(x_i, q_i), \quad x = (x_1, \dots, x_n), \quad q \in \mathcal{P}(\Omega_n),$$

*where as usual  $q_i \in \mathcal{P}(\Omega_1)$  denotes the  $i$ -th marginal of  $q$ . Then, given  $\hat{q} \in \mathcal{P}(I_n)$ , for all  $\ell \in I_n$  and all  $x, y \in \Omega_n^\ell$ , it holds  $\inf_q c^{(n)}(x, q) = \inf_q c^{(n)}(y, q)$ , where the infimum runs over all  $q \in \mathcal{P}(\Omega_n)$  so that  $q(\Omega_n^k) = \hat{q}(k)$  for all  $k \in I_n$ . In particular, one can define a cost function on  $I_n$ ,  $\hat{c}: I_n \times \mathcal{P}(I_n) \rightarrow [0, \infty]$ , by  $\hat{c}(\ell, \hat{p}) := \inf_q c(x, q)$  where  $x \in \Omega_n^\ell$  is arbitrary. Moreover, such a cost satisfies the following properties:*

- (i) *If  $q \mapsto c(x, q)$  is convex for all  $x$ , then so is  $\hat{q} \mapsto \hat{c}(\ell, \hat{q})$  for all  $n$  and all  $\ell \in I_n$ .*
- (ii) *If  $c(x, q) = \theta \left( x - \int_{\Omega_1} y q(dy) \right)$  (on  $\Omega_1$ ) for some convex function  $\theta: \mathbb{R} \rightarrow \mathbb{R}_+$ , then it holds*

$$\hat{c}(\ell, \hat{q}) \geq n\theta \left( \frac{1}{n} \left( \ell - \int_{I_n} k \hat{q}(dk) \right) \right), \quad \forall \ell \in I_n, \quad \forall \hat{q} \in \mathcal{P}(I_n).$$

- (iii) *If  $c(x, q) = \theta \left( \int_{\Omega_1} |x - y| q(dy) \right)$  (on  $\Omega_1$ ) for some convex function  $\theta: \mathbb{R} \rightarrow \mathbb{R}_+$ , then it holds*

$$\hat{c}(\ell, \hat{q}) \geq n\theta \left( \frac{1}{n} \left( \int_{I_n} |\ell - k| \hat{q}(dk) \right) \right), \quad \forall \ell \in I_n, \quad \forall \hat{q} \in \mathcal{P}(I_n).$$

(iv) Assume that there exist  $a_1, a_2 \geq 0$  such that it holds

$$\mathcal{T}_{c^{(n)}}(\nu_1|\nu_2) \leq a_1 H(\nu_1|\mu_\rho^n) + a_2 H(\nu_2|\mu_\rho^n), \quad \forall \nu_1, \nu_2 \in \mathcal{P}(\Omega_n).$$

Then, it holds

$$\mathcal{T}_{\hat{c}}(\hat{\nu}_1|\hat{\nu}_2) \leq a_1 H(\hat{\nu}_1|\mu_{n,\rho}) + a_2 H(\hat{\nu}_2|\mu_{n,\rho}), \quad \forall \hat{\nu}_1, \hat{\nu}_2 \in \mathcal{P}(I_n).$$

*Proof.* Fix  $\ell \in I_n$ ,  $x, y \in \Omega_n^\ell$  and  $\hat{q} \in \mathcal{P}(I_n)$ . Then, since  $x, y \in \Omega_n^\ell$  have the same number of ones and zeros, there exists a permutation  $\sigma \in \mathcal{S}_n$  so that  $y_i = x_{\sigma(i)}$  for all  $i$ . Given  $q \in \mathcal{P}(\Omega_n)$  satisfying  $q(\Omega_n^k) = \hat{q}(k)$  for all  $k \in I_n$ , define  $q_\sigma \in \mathcal{P}(\Omega_n)$  by  $q_\sigma(z) = q(z_\sigma)$  where we set for simplicity  $z_\sigma := (z_{\sigma^{-1}(1)}, \dots, z_{\sigma^{-1}(n)})$  (with  $\sigma^{-1}$  being the inverse of  $\sigma$ ). It is easy to verify that (1) :  $q_\sigma \in \mathcal{P}(\Omega_n)$  is such that  $q_\sigma(\Omega_n^k) = \hat{q}(k)$  for all  $k \in I_n$ , (2) : the  $i$ -th marginal of  $q_\sigma$  equal the  $\sigma(i)$ -th marginal of  $q$ :  $(q_\sigma)_i = q_{\sigma(i)}$ . Hence  $c^{(n)}(y, q_\sigma) = \sum_{i=1}^n c(y_i, (q_\sigma)_i) = \sum_{i=1}^n c(x_{\sigma(i)}, q_{\sigma(i)}) = \sum_{i=1}^n c(x_i, q_i)$ . In turn  $\inf_q c^{(n)}(y, q) \leq \inf_q c^{(n)}(x, q)$  and the first part of the lemma follows.

Item (i) is easy to verify and follows from the fact that the constraints on  $q$  are linear.

As for Item (ii), fix  $\ell \in I_n$  and  $\hat{q} \in \mathcal{P}(I_n)$ . Then, for all  $x \in \Omega_n^\ell$  and all  $q \in \mathcal{P}(\Omega_n)$  such that  $q(\Omega_n^k) = \hat{q}(k)$  for all  $k \in I_n$ , by convexity it holds

$$\begin{aligned} c^{(n)}(x, p) &= \sum_{i=1}^n \theta \left( x_i - \int_{\Omega_1} y_i q_i(dy_i) \right) \geq n\theta \left( \frac{1}{n} \sum_{i=1}^n \left( x_i - \int_{\Omega_1} y_i q_i(dy_i) \right) \right) \\ &= n\theta \left( \frac{1}{n} \left( \sum_{i=1}^n x_i - \int_{\Omega_n} \left( \sum_{i=1}^n y_i \right) q(dy) \right) \right) = n\theta \left( \frac{1}{n} \left( \ell - \int_{I_n} k d\hat{q}(k) \right) \right). \end{aligned}$$

Taking the infimum over all  $q \in \mathcal{P}(\Omega_n)$  such that  $q(\Omega_n^k) = \hat{q}(k)$  for all  $k \in I_n$  yields the desired result.

The proof of Item (iii) is similar and left to the reader (use the triangle inequality).

To prove Item (iv), fix  $\hat{\nu}_1, \hat{\nu}_2 \in \mathcal{P}(I_n)$  and define  $\nu_1 \in \mathcal{P}(\Omega_n)$  by  $\nu_1(x) = \hat{\nu}_1(\varphi(x))/(\varphi(x))$ ,  $x \in \Omega_n$ , where we recall that  $\varphi$  denotes the projection  $\varphi(x) = \sum_{i=1}^n x_i$ . Then,  $H(\nu_1|\mu_\rho^n) = H(\hat{\nu}_1|\mu_{n,\rho})$ . Hence, defining identically  $\nu_2 \in \mathcal{P}(\Omega_n)$  from  $\hat{\nu}_2$ , the result follows if we prove that  $\mathcal{T}_{c^{(n)}}(\nu_1|\nu_2) \geq \mathcal{T}_{\hat{c}}(\hat{\nu}_1|\hat{\nu}_2)$ . By Theorem 3.5, and restricting the supremum by using the projection  $\varphi$ , we have

$$\begin{aligned} \mathcal{T}_{c^{(n)}}(\nu_1|\nu_2) &= \sup_{\Psi \in \Phi(\Omega_n)} \left\{ \int_{\Omega_n} \left( \inf_{q \in \mathcal{P}(\Omega_n)} \int_{\Omega_n} \Psi(y) q(dy) + c^{(n)}(x, q) \right) \nu_2(dx) - \int_{\Omega_n} \Psi(y) \nu_1(dy) \right\} \\ &\geq \sup_{\hat{\Psi} \in \Phi(I_n)} \left\{ \int_{\Omega_n} \left( \inf_{q \in \mathcal{P}(\Omega_n)} \int_{\Omega_n} \hat{\Psi}(\varphi(y)) q(dy) + c^{(n)}(x, q) \right) \nu_2(dx) - \int_{\Omega_n} \hat{\Psi}(\varphi(y)) \nu_1(dy) \right\}. \end{aligned}$$

Hence, by the first part of the lemma and again Theorem 3.5, we have

$$\begin{aligned} &\mathcal{T}_{c^{(n)}}(\nu_1|\nu_2) \\ &\geq \sup_{\hat{\Psi} \in \Phi(I_n)} \left\{ \int_{\Omega_n} \inf_{\hat{q} \in \mathcal{P}(I_n)} \inf_{\substack{q \in \mathcal{P}(\Omega_n): \\ q(\Omega_n^k) = \hat{q}(k), \forall k \in I_n}} \left( \int_{\Omega_n} \hat{\Psi}(\varphi(y)) q(dy) + c^{(n)}(x, q) \right) \nu_2(dx) - \int_{I_n} \hat{\Psi}(k) \hat{\nu}_1(dk) \right\} \\ &= \sup_{\hat{\Psi} \in \Phi(I_n)} \left\{ \int_{\Omega_n} \left( \inf_{\hat{q} \in \mathcal{P}(I_n)} \int_{I_n} \hat{\Psi}(k) \hat{q}(dk) + \hat{c}(\varphi(x), \hat{q}) \right) \nu_2(dx) - \int_{I_n} \hat{\Psi}(k) \hat{\nu}_1(dk) \right\} \\ &= \sup_{\hat{\Psi} \in \Phi(I_n)} \left\{ \int_{I_n} \left( \inf_{\hat{q} \in \mathcal{P}(I_n)} \int_{I_n} \hat{\Psi}(k) \hat{q}(dk) + \hat{c}(k, \hat{q}) \right) \hat{\nu}_2(dk) - \int_{I_n} \hat{\Psi}(k) \hat{\nu}_1(dk) \right\} = \mathcal{T}_{\hat{c}}(\hat{\nu}_1|\hat{\nu}_2). \end{aligned}$$



This ends the proof of the lemma.  $\square$

As a consequence of the above lemma we get the following weak transport inequalities for the binomial distribution  $\mu_{n,\rho}$ .

**Corollary 7.7.** *For all  $t \in (0, 1)$ , all  $\rho \in (0, 1)$  and all  $n = 1, 2, \dots$ , it holds*

$$(7.8) \quad \overline{\mathcal{T}}_{\theta_{\rho,t,n}}(\nu_1|\nu_2) \leq \frac{1}{1-t}H(\nu_1|\mu_{n,\rho}) + \frac{1}{t}H(\nu_2|\mu_{n,\rho}), \quad \forall \nu_1, \nu_2 \in \mathcal{P}(I_n).$$

In particular

$$\overline{\mathcal{T}}_{\theta_{\rho,1,n}}(\mu_{n,\rho}|\nu) \leq H(\nu|\mu_{n,\rho}) \quad \text{and} \quad \overline{\mathcal{T}}_{\theta_{\rho,0,n}}(\nu|\mu_{n,\rho}) \leq H(\nu|\mu_{n,\rho}), \quad \forall \nu \in \mathcal{P}(I_n)$$

where  $\theta_{\rho,t,n}(h) := n\theta_{\rho,t}(h/n)$ ,  $h \in \mathbb{R}$ .

Moreover, the same inequalities hold replacing  $\overline{\mathcal{T}}_{\theta_{\rho,t,n}}$ ,  $\overline{\mathcal{T}}_{\theta_{\rho,1,n}}$  and  $\overline{\mathcal{T}}_{\theta_{\rho,0,n}}$  respectively by  $\tilde{\mathcal{T}}_{\tilde{\theta}_{\rho,t,n}}$ ,  $\tilde{\mathcal{T}}_{\tilde{\theta}_{\rho,1,n}}$  and  $\tilde{\mathcal{T}}_{\tilde{\theta}_{\rho,0,n}}$ , where  $\tilde{\theta}_{\rho,t,n}(h) := \min(\theta_{\rho,t,n}(h), \theta_{\rho,t,n}(-h))$ ,  $h \in \mathbb{R}$ .

*Proof.* The inequalities involving  $\overline{\mathcal{T}}$  follow easily from Lemma 7.6 (Point (ii) and (iv)) and Corollary 7.4. Similarly the inequalities involving  $\tilde{\mathcal{T}}$  follow from Lemma 7.6 (Point (iii) and (iv)) and Corollary 7.4, once one shows that  $\tilde{\theta}_{\rho,t,n}$  is convex. This is a simple consequence of the fact that  $\tilde{\theta}_{\rho,t,n}(h) = \theta_{\rho \wedge 1 - \rho, t, n}(|h|)$  (see the beginning of the section) and that  $\theta_{\rho,t}$  is convex as proved in Appendix A. This completes the proof.  $\square$

**7.3. Weak transport cost inequality for the Poisson measure.** In this section we derive a weak transport-entropy inequality for the Poisson probability measure  $p_\lambda$ , with parameter  $\lambda > 0$ : for all  $k \in \mathbb{N}$ ,  $p_\lambda(k) = \frac{\lambda^k}{k!}e^{-\lambda}$ . The idea is to use the weak convergence of the binomial distribution  $\mu_{n,\rho_n}$ , with  $\rho_n := \lambda/n$ , towards the Poisson measure  $p_\lambda$ .

Set, for  $t \in (0, 1)$ ,  $h \in \mathbb{R}$ ,

$$c_{\lambda,t}(h) := \lim_{n \rightarrow \infty} n\theta_{\rho_n,t}\left(\frac{h}{n}\right).$$

The convexity of  $\theta_{\rho_n,t}$  provides the convexity of the cost function  $c_{\lambda,t}$ . We claim that

$$(7.9) \quad c_{\lambda,t}(h) = \left[ \frac{\lambda}{t} w\left(\frac{r_t(h)}{\lambda}\right) + \frac{\lambda}{1-t} w\left(\frac{h + r_t(h)}{\lambda}\right) \right] \mathbf{1}_{h \leq 0},$$

where  $r = r_t(h) \in [0, \lambda)$  is the unique solution of the following equation,

$$(7.10) \quad (\lambda - r)^{1-t}(\lambda - r - h)^t = \lambda, \quad h \leq 0.$$

The technical proof of this claim is given in Appendix A. We may observe that  $\lim_{t \rightarrow 0} r_t(h) = 0$  and  $\lim_{t \rightarrow 1} r_t(h) = \min(-h, \lambda)$  (see the end of the proof of Proposition 7.11 below).

Set also

$$\overline{\mathcal{Q}}_{c_{\lambda,t}}f(\ell) = \inf_{q \in \mathcal{P}(\mathbb{N})} \left\{ \int f dq + c_{\lambda,t} \left( \ell - \int k q(dk) \right) \right\} \quad \ell \in \mathbb{N},$$

the corresponding inf-convolution operators (for all  $f: \mathbb{N} \rightarrow \mathbb{R}$ , say bounded).

**Proposition 7.11.** *For all  $\lambda > 0$ ,  $t \in (0, 1)$ , it holds*

$$(7.12) \quad \overline{\mathcal{T}}_{c_{\lambda,t}}(\nu_1|\nu_2) \leq \frac{1}{1-t}H(\nu_1|p_\lambda) + \frac{1}{t}H(\nu_2|p_\lambda), \quad \forall \nu_1, \nu_2 \in \mathcal{P}(\mathbb{N}).$$

we also have

$$(7.13) \quad \overline{\mathcal{T}}_{c_{\lambda,0}}(p_\lambda|\nu) \leq H(\nu|p_\lambda), \quad \forall \nu \in \mathcal{P}(\mathbb{N}),$$

with  $c_{\lambda,0}(h) := \lambda w\left(\frac{h}{\lambda}\right) \mathbf{1}_{h \leq 0}$ , and

$$(7.14) \quad \overline{T}_{c_{\lambda,1}}(\nu|p_\lambda) \leq H(\nu|p_\lambda), \quad \forall \nu \in \mathcal{P}(\mathbb{N}),$$

with  $c_{\lambda,1}(h) := \lambda w\left(\frac{-h}{\lambda}\right) \mathbf{1}_{h \leq 0}$ .

The weak transport inequalities (7.13) and (7.14) are the boundary cases of the weak transport inequality (7.12) when  $t$  goes to 0 or to 1 and  $\nu_2 = p_\lambda$  or  $\nu_1 = p_\lambda$ .

**Remark 7.15.** From the proof it will be clear that our approach fails to give a weak transport-entropy inequality involving  $\tilde{T}$  for the Poisson measure  $p_\lambda$ . Indeed, one of the key ingredients is to use the limit  $\lim_{n \rightarrow \infty} n\theta(h/n)$  which, in the case of the  $\tilde{T}$  cost, is trivial:  $\lim_{n \rightarrow \infty} n\tilde{\theta}(h/n) = 0$  for all  $h \in \mathbb{R}$ . However there is no clear evidence that such inequalities do not hold.

Finally, we observe that (7.14) and (7.13) are optimal, i.e. the constant 1 cannot be improved. Indeed, e.g. (7.14) is equivalent, thanks to Proposition 5.5, to

$$\exp\left\{\int \overline{Q}_{c_{\lambda,0}} f dp_\lambda\right\} \int e^{-f} dp_\lambda \leq 1,$$

which is an equality for  $f(x) = -tx$ ,  $x \in \mathbb{R}$ ,  $t \geq 0$  (the same holds for (7.13)).

*Proof.* We first start with the proof of (7.12). Recall that  $\mu_{n,\rho}$  denotes the binomial distribution on  $I_n = \{0, 1, \dots, n\}$ . From Proposition 5.5 (i.e. the Bobkov-Götze dual characterization) and Corollary 7.7, for all integers  $n$ , all  $\rho \in (0, 1)$  and all bounded function  $f$  on  $\mathbb{N}$ , it holds

$$(7.16) \quad \left(\int e^{t\overline{Q}_{\theta_{\rho,t,n}} f} d\mu_{n,\rho}\right)^{1/t} \left(\int e^{-(1-t)f} d\mu_{n,\rho}\right)^{1/(1-t)} \leq 1.$$

where we recall that  $\overline{Q}_{\theta_{\rho,t,n}} f$  is the (bar) infimum convolution of  $f$  associated to the cost function  $n\theta_{\rho,t}(\cdot/n)$  defined by

$$\overline{Q}_{\theta_{\rho,t,n}} f(\ell) = \inf_{q \in \mathcal{P}(I_n)} \left\{ \int_{I_n} f dq + n\theta_{\rho,t} \left( \frac{1}{n} \left( \ell - \int_{I_n} k q(dk) \right) \right) \right\}, \quad \ell \in I_n.$$

Our aim is to take the limit in (7.16), with  $\rho = \rho_n := \lambda/n$ . To achieve it, we need to prove that inverting infimum and limit goes in the right direction. Namely we shall prove the following claim.

**Claim 7.17.** *It holds*

$$\lim_{n \rightarrow \infty} \overline{Q}_{\theta_{\rho_n,t,n}} f(\ell) \geq \overline{Q}_{c_{\lambda,t}} f(\ell), \quad \ell \in \mathbb{N}.$$

We postpone the proof of the claim for a moment, but first show using it to complete the proof of (7.12). Using the claim, we get from (7.16), by the weak convergence of  $\mu_{n,\rho_n}$  towards  $p_\lambda$  as  $n$  goes to  $\infty$ ,

$$(7.18) \quad \left(\int e^{t\overline{Q}_{c_{\lambda,t}} f} dp_\lambda\right)^{1/t} \left(\int e^{-(1-t)f} dp_\lambda\right)^{1/(1-t)} \leq 1.$$

The thesis then follows by the Bobkov-Götze dual characterization (Proposition 5.5).

Now it remains to prove Claim 7.17. Set  $\|f\|_\infty := \sup_{k \in \mathbb{N}} |f(k)|$ . Since  $\theta_{\rho_n,t} \geq 0$ , and  $\theta_{\rho_n,t}(0) = 0$  it holds

$$\overline{Q}_{\theta_{\rho_n,t,n}} f(\ell) \geq \inf_{q \in \mathcal{P}(I_n)} \left\{ \int_{I_n} f dq + n\theta_{\rho_n,t} \left( -\frac{1}{n} \left[ \ell - \int_{I_n} k q(dk) \right]_- \right) \right\}, \quad \ell \in I_n$$

where  $[X]_- := \max(-X, 0)$  denotes the negative part. The above infimum is reached by compactness at some  $\hat{q}$  (that depends on  $\rho, t, n, k$ ) satisfying:

$$\int_{I_n} f d\hat{q} + n\theta_{\rho_n, t} \left( -\frac{1}{n} \left[ \ell - \int_{I_n} k \hat{q}(dk) \right]_- \right) \leq \overline{Q}_{\theta_{\rho_n, t, n}} f(\ell) \leq \|f\|_\infty.$$

At this point we claim that for all  $t \in (0, 1)$ ,  $h \geq 0$  and  $n \geq 2\lambda$ ,

$$(7.19) \quad n\theta_{\rho_n, t}(-h/n) \geq \frac{1}{1-t} w \left( -(1-t) \frac{h}{2\lambda} \right) := v_t(h).$$

The proof of this claim is given in Appendix A. Let  $v_t^{-1}$  denote the inverse function of increasing bijection  $v_t : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ . It follows that

$$\left[ \ell - \int k \hat{q}(dk) \right]_- \leq v_t^{-1}(2\|f\|_\infty), \quad \forall \ell \in I_n$$

In turn, since  $\mathcal{P}(I_n) \subset \mathcal{P}(\mathbb{N})$ ,

$$\begin{aligned} \overline{Q}_{\theta_{\rho_n, t, n}} f(\ell) &\geq \int f d\hat{q} + c_{\lambda, t} \left( \ell - \int k \hat{q}(dk) \right) - \sup_{-v_t^{-1}(2\|f\|_\infty) \leq h \leq 0} |n\theta_{\rho_n, t}(h/n) - c_{\lambda, t}(h)| \\ &\geq \overline{Q}_{c_{\lambda, t}} f(\ell) - \sup_{-v_t^{-1}(2\|f\|_\infty) \leq h \leq 0} |n\theta_{\rho_n, t}(h/n) - c_{\lambda, t}(h)|. \end{aligned}$$

The pointwise convergence of  $n\theta_{\rho_n, t}(h/n)$  to  $c_{\lambda, t}(h)$  and the monotonicity of  $\theta_{\rho_n, t}$  on  $\mathbb{R}_-$  implies, according to a classical variant of Dini's theorem, that

$$\lim_{n \rightarrow \infty} \sup_{-v_t^{-1}(2\|f\|_\infty) \leq h \leq 0} |n\theta_{\rho_n, t}(h/n) - c_{\lambda, t}(h)| = 0.$$

The proof of the weak transport inequality (7.12) is completed.

The boundary cases (7.13) and (7.14) of Proposition 7.11 can similarly be obtained from the Bobkov-Götze dual characterization of (the boundary cases of) weak transport inequalities of Corollary 7.7. For  $t = 0$ , the proof to get (7.13) is identical, replacing  $t$  by 0. In the proof of (7.14) for  $t = 1$ , we need to replace the claim (7.19) by the following inequality: for all  $h \geq 0$ ,

$$n\theta_{\rho_n, 0}(-h/n) \geq h^2/2\lambda.$$

This easy computation and the details of the proof of (7.13) and (7.14) are left to the reader.

We could also obtain (7.13) and (7.14) from (7.18) as  $t$  goes to 0 or to 1, with more computations for appropriate justifications. Let us just show that for all  $h \geq 0$ ,

$$\lim_{t \rightarrow 0} c_{\lambda, t}(-h) = \lambda w \left( \frac{-h}{\lambda} \right) = c_{\lambda, 0}(-h),$$

and for all  $h \geq 0$ ,  $h \neq 1$ ,

$$\lim_{t \rightarrow 1} c_{\lambda, t}(-h) = \lambda w \left( \frac{h}{\lambda} \right) = c_{\lambda, 1}(-h).$$

Let  $\tilde{h} = h/\lambda$  and  $\tilde{r}(h) = \tilde{r}_t(h) = r_t(-h)/\lambda \leq 1$ . Rewriting (7.10), one has for all  $h \geq 0$ ,

$$(1-t) \log(1 - \tilde{r}(h)) + t \log(1 - h - \tilde{r}(h)) = 0.$$

This implies that

$$\frac{\tilde{h} + 1 - \sqrt{(\tilde{h} + 1)^2 - 4t\tilde{h}}}{2} \leq \tilde{r}(h) \leq \min(t\tilde{h}, 1).$$

The second inequality is obtained by applying twice the inequality  $\log(1-u) \leq u$ ,  $u < 1$ , and the first inequality by applying twice the inequality  $\log(1-u) \geq -u/(1-u)$ ,  $u < 1$  followed by few easy computations.

These inequalities ensure that  $\lim_{t \rightarrow 0} \tilde{r}_t(h) = 0$  and  $\lim_{t \rightarrow 1} \tilde{r}_t(h) = \min(1, \tilde{h})$ . Then the given limits of  $c_{\lambda,t}(-h)$ , as  $t$  goes to 0 or to 1, easily follow.  $\square$

**7.4. Sufficient condition for  $\overline{\mathbf{T}}_2^-$  on the line.** In this short section, we would like to take advantage of some known results from [1] to give a sufficient condition for the transport-entropy inequality  $\overline{\mathbf{T}}_2^-$  to hold on the line.

Our starting point is the following result.

**Theorem 7.20** ([1]). *Let  $\mu$  be a symmetric probability measure on the line. Assume that there exists  $c > 0$  and  $\alpha < 1$  such that for all  $x \geq 0$ ,  $\mu([x + \frac{c}{x}, \infty)) \leq \alpha\mu([x, \infty))$ . Then, there exists  $C(c, \alpha) \in (0, \infty)$  such that for every smooth convex function  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$ , it holds*

$$\text{Ent}_\mu(e^\varphi) \leq C(c, \alpha) \int \varphi'^2 e^\varphi d\mu.$$

Observe that we assumed symmetry for simplicity. It is not essential and a similar result holds for non symmetric measures.

**Corollary 7.21.** *Let  $\mu$  be a symmetric probability measure on the line. Assume that there exists  $c > 0$  and  $\alpha < 1$  such that for all  $x \geq 0$ ,  $\mu([x + \frac{c}{x}, \infty)) \leq \alpha\mu([x, \infty))$ . Then, there exists  $C = C(c, \alpha) \in (0, \infty)$  such that  $\overline{\mathbf{T}}_2^-(C)$  holds.*

*Proof.* Theorem 7.20 guarantees that Item (iii) of Theorem 6.8 holds, with  $1/(2\rho) = C(c, \alpha)$  (Choose  $\|\cdot\| = |\cdot|$ , where  $|\cdot|$  is the absolute value, so that  $\|\cdot\|_* = |\cdot|$ ). The desired result follows from Theorem 6.8.  $\square$

**7.5. Weak transport cost inequalities with the Hamming distance.** In this section, we recall some known universal transport-entropy inequalities associated to the weak transport costs  $\tilde{\mathcal{T}}$  and  $\hat{\mathcal{T}}$ . The following results are partially given in [31]. We start with some notations. For  $t \in (0, 1)$ , let  $\alpha_t(x, p) = \tilde{\alpha}_t(\int \mathbf{1}_{x \neq y} p(dy))$ ,  $x \in X$ ,  $p \in \mathcal{P}(X)$ , with

$$\tilde{\alpha}_t(u) = \begin{cases} \frac{t(1-u)\log(1-u) - (1-tu)\log(1-tu)}{t(1-t)} & \text{if } 0 \leq u \leq 1 \\ +\infty & \text{otherwise.} \end{cases}$$

Also, let  $\beta_t(x, p) = \int \hat{\beta}_t(\mathbf{1}_{x \neq y} \frac{dp}{d\mu}(y)) \mu(dy)$ ,  $x \in X$ , if  $p \in \mathcal{P}(X)$  is absolutely continuous with respect to  $\mu$  on  $X \setminus \{x\}$ , and  $\beta_t(x, p) = +\infty$  otherwise, with for all  $u \geq 0$ ,  $\hat{\beta}_t(u) := \sup_{s \in \mathbb{R}} \{su - \hat{\beta}_t^*(s)\}$  and  $\hat{\beta}_t^*(s) := \frac{te^{(1-t)s} + (1-t)e^{-ts} - 1}{t(1-t)}$ ,  $s \in \mathbb{R}$ .

**Proposition 7.22.** *Let  $(X, d)$  be a compact metric space or a countable set of isolated points. Let  $t \in (0, 1)$  and  $\mu \in \mathcal{P}(X)$ . Then,*

(1) *For all probability measures  $\nu_1, \nu_2$  on  $X$ , it holds*

$$\tilde{\mathcal{T}}_{\alpha_t}(\nu_1 | \nu_2) \leq \frac{1}{1-t} H(\nu_1 | \mu) + \frac{1}{t} H(\nu_2 | \mu).$$

(2) *For all probability measures  $\nu_1, \nu_2$  on  $X$  absolutely continuous with respect to  $\mu$ , it holds*

$$\hat{\mathcal{T}}_{\beta_t}(\nu_1 | \nu_2) \leq \frac{1}{1-t} H(\nu_1 | \mu) + \frac{1}{t} H(\nu_2 | \mu).$$

A short proof of the first point of this proposition is given in [31] (see Lemma 2.1.). By Proposition 5.5 and Corollary 3.12, the second point of this proposition is exactly the dual form of Theorem 1.1 of [31] (for  $n = 1$ ). This new expected formulation of Theorem 1.1

in [31] is therefore a direct consequence of the generalization of the Kantorovich theorem (Theorem 3.5).

Let us just recall one simple and useful corollary of Proposition 7.22 for probability measures with bounded support.

**Corollary 7.23.** *Suppose that  $\mu$  is a probability on  $\mathbb{R}^m$  (equipped with some arbitrary norm  $\|\cdot\|$ ) such that the diameter of  $\text{supp}(\mu)$  is bounded by  $M > 0$ . Then  $\mu$  satisfies the inequality  $\tilde{\mathbf{T}}_2(4M^2, 4M^2)$  and thus  $\overline{\mathbf{T}}_2(4M^2, 4M^2)$ .*

*Proof.* Observe that  $\tilde{\alpha}_t(u) \geq u^2/2$ , for all  $u \in [0, 1]$  and  $t = 1/2$ . Furthermore, if  $\nu_1, \nu_2$  are absolutely continuous with respect to  $\mu$  then  $\text{supp}(\nu_i) \subset \text{supp}(\mu)$ . Therefore, if  $\pi(dx dy) = \nu_1(dx)p_x(dy)$  is a coupling between  $\nu_1$  and  $\nu_2$ , then  $\int \|x - y\| p_x(dy) \leq M \int \mathbf{1}_{\{x \neq y\}} p_x(dy)$  for  $\nu_1$  almost all  $x$ , and so

$$\begin{aligned} \frac{1}{2M^2} \int \left( \int \|x - y\| p_x(dy) \right)^2 \nu_1(dx) &\leq \int \tilde{\alpha}_t \left( \frac{1}{M} \int \|x - y\| p_x(dy) \right) \nu_1(dx) \\ &\leq \int \tilde{\alpha}_t \left( \int \mathbf{1}_{\{x \neq y\}} p_x(dy) \right) \nu_1(dx). \end{aligned}$$

Optimizing over all  $\pi$  and then using Proposition 7.22 for  $t = 1/2$  completes the proof.  $\square$

## 8. PROOF OF THE DUALITY THEOREM AND OF ITS COROLLARIES

This section is dedicated to the proofs of Theorem 3.5 and Corollaries 3.7, 3.9, 3.11 and 3.12.

The proof of Theorem 3.5 uses classical tools from convex analysis that we recall in a separate subsection (see Section 8.2 below), and then apply them to our specific setting. We refer to Mikami [25], Léonard [19], Tan-Touzi [35] for similar strategies.

**8.1. Fenchel-Legendre duality.** The main tool used in the proof of Theorem 3.5 is the following Fenchel-Legendre duality theorem (see for instance [39, Theorem 2.3.3]).

**Theorem 8.1** (Fenchel-Legendre duality theorem). *Let  $E$  be a Hausdorff locally convex topological vector space and  $E'$  its topological dual space. For any lower semicontinuous convex function  $F: E \rightarrow ]-\infty, \infty]$ , it holds*

$$F(x) = \sup_{\ell \in E'} \{\ell(x) - F^*(\ell)\}, \quad x \in E,$$

where the Fenchel-Legendre transform  $F^*$  of  $F$  is defined by

$$F^*(\ell) = \sup_{x \in E} \{\ell(x) - F(x)\}, \quad \ell \in E'.$$

To apply Theorem 8.1 in our framework, one needs to identify the topological dual space of  $\mathcal{M}_\gamma(X)$  equipped with the topology  $\sigma(\mathcal{M}_\gamma(X))$  defined in Section 2. More precisely, the next lemma will enable us to identify the dual space  $(\mathcal{M}_\gamma(X), \sigma(\mathcal{M}_\gamma(X)))'$  to the set  $\Phi_\gamma(X)$ .

**Lemma 8.2.** *A linear form  $\ell: \mathcal{M}_\gamma(X) \rightarrow \mathbb{R}$  is continuous with respect to the topology  $\sigma(\mathcal{M}_\gamma(X))$  if and only if there exists  $\varphi \in \Phi_\gamma(X)$  such that*

$$\ell(m) = \int \varphi dm, \quad \forall m \in \mathcal{M}_\gamma(X).$$

The proof of this lemma appears, for instance, in the book by Deuschel and Stroock [8]. We recall it here for the sake of completeness.

*Proof of Lemma 8.2.* The fact that linear functionals of the form  $m \mapsto \int \varphi dm$ ,  $\varphi \in \Phi_\gamma$  are continuous comes from the very definition of the topology  $\sigma(\mathcal{M}_\gamma(X))$ . Conversely, let  $\ell$  be a continuous linear functional and let us show that  $\ell$  is of the preceding form. Define  $\varphi(x) = \ell(\delta_x)$ ,  $x \in X$  (where  $\delta_x$  is the Dirac mass at  $x$ ). First we will show that  $\varphi$  belongs to  $\Phi_\gamma(X)$ . The map  $X \ni x \mapsto \delta_x \in \mathcal{M}_\gamma(X)$  is continuous. Namely, for all  $\varphi_1, \dots, \varphi_n \in \Phi_\gamma$ , it holds  $\{x \in X; \delta_x \in \cap_{i=1}^n U_{\varphi_i, a_i, \varepsilon_i}\} = \{x \in X; |\varphi_i(x) - a_i| < \varepsilon_i, \forall i \leq n\}$ , (where  $U_{\varphi_i, a_i, \varepsilon_i}$  is defined by (2.3)) and this set is open, which proves that  $x \mapsto \delta_x$  is continuous on  $X$ . As a result  $\varphi$  is continuous. It remains to prove that  $\varphi$  satisfies the growth condition (2.2). Since  $\ell$  is continuous, the set  $O := \{m \in \mathcal{M}_\gamma(X); |\ell(m)| < 1\}$  is open and contains 0. By definition of the topology  $\sigma(\mathcal{M}_\gamma(X))$ , there exist an integer  $n$ ,  $\varphi_1, \dots, \varphi_n \in \Phi_\gamma$ ,  $a_1, \dots, a_n \in \mathbb{R}$  and  $\varepsilon_1, \dots, \varepsilon_n > 0$  such that  $O$  contains  $\cap_{i=1}^n U_{\varphi_i, a_i, \varepsilon_i}$  and  $0 \in \cap_{i=1}^n U_{\varphi_i, a_i, \varepsilon_i}$ . As a result,

$$0 \in \bigcap_{i=1}^n U_{\varphi_i, a_i, \varepsilon_i} \quad \Rightarrow \quad A := \max_{i \in \{1, \dots, n\}} \left| \frac{a_i}{\varepsilon_i} \right| < 1,$$

and (given  $m \in \mathcal{M}_\gamma(X)$ )

$$\sum_{i=1}^n \left| \int \frac{\varphi_i}{\varepsilon_i} dm \right| < 1 - A \quad \Rightarrow \quad m \in O.$$

Thus, since  $m/\ell(m) \notin O$ ,

$$|\ell(m)| \leq \frac{1}{1-A} \sum_{i=1}^n \left| \int \frac{\varphi_i}{\varepsilon_i} dm \right|, \quad \forall m \in \mathcal{M}_\gamma(X).$$

Applying this inequality to  $m = \delta_x$  and using the growth conditions (2.2) satisfied by the  $\varphi_i$ 's, one sees that  $\varphi$  verifies (2.2).

Finally, let us show that  $\ell(m) = \int \varphi dm$ , for all  $m \in \mathcal{M}_\gamma(X)$ . If  $m$  is a linear combination of Dirac measures, then this identity is clearly satisfied. Since any measure  $m$  can be approached in the topology  $\sigma(\mathcal{M}_\gamma(X))$  by some sequence  $m_n$  of measures with finite support, the equality  $\ell(m) = \int \varphi dm$  extends to any  $m \in \mathcal{M}_\gamma(X)$ .  $\square$

During the proof of Theorem 3.5, we will also use the following easy extension of Prokhorov theorem.

**Theorem 8.3.** *A set  $A \subset \mathcal{P}_\gamma(X)$  is relatively compact for the topology  $\sigma(\mathcal{M}_\gamma(X))$  if and only if for all  $\varepsilon > 0$ , there exists a compact set  $K_\varepsilon \subset X$  such that*

$$\int_{X \setminus K_\varepsilon} (1 + \gamma(d(x_o, x))) \nu(dx) \leq \varepsilon, \quad \forall \nu \in A,$$

where  $x_o$  is some arbitrary fixed point.

## 8.2. Proof of Theorem 3.5 (Duality).

*Proof of Theorem 3.5.* Fix  $\mu \in \mathcal{P}_\gamma(X)$  and let us consider the function  $F$  defined on  $\mathcal{M}_\gamma(X)$  by

$$F(m) = \mathcal{T}_c(m|\mu), \text{ if } m \in \mathcal{P}_\gamma(X) \quad \text{and} \quad F(m) = +\infty \text{ otherwise.}$$

Let us show that the function  $F$  satisfies the assumptions of Theorem 8.1.

First we will prove that  $F$  is convex on  $\mathcal{M}_\gamma(X)$ . According to the definition of  $F$ , it is clearly enough to prove the convexity of  $F$  over (the convex set)  $\mathcal{P}_\gamma(X)$ . Take  $\nu_0, \nu_1 \in \mathcal{P}_\gamma(X)$  and  $\pi_i \in \Pi(\mu, \nu_i)$   $i = 0, 1$  with disintegration kernels  $(p_x^0)_{x \in X}, (p_x^1)_{x \in X}$ . Then for all  $t \in [0, 1]$ ,  $\pi_t := (1-t)\pi_0 + t\pi_1 \in \Pi(\mu, (1-t)\nu_0 + t\nu_1)$  and its disintegration

kernel satisfies  $p_x^t = (1-t)p_x^0 + tp_x^1$ , for  $\mu$  almost every  $x \in X$ . Since the cost function  $c$  is convex in its second argument, it holds

$$F((1-t)\nu_0 + t\nu_1) \leq I_c[\pi_t] = \int c(x, p_x^t) \mu(dx) \leq (1-t)I_c[\pi_0] + tI_c[\pi_1].$$

Optimizing over  $\pi_0, \pi_1$  gives  $F((1-t)\nu_0 + t\nu_1) \leq (1-t)F(\nu_0) + tF(\nu_1)$ , which proves the desired convexity property.

Next we will prove that  $F$  is lower-semicontinuous, for the topology  $\sigma(\mathcal{M}_\gamma(X))$ , on  $\mathcal{M}_\gamma(X)$ . Let  $(m_n)_n$  be a sequence of  $\mathcal{M}_\gamma(X)$  converging to some  $m$ . One needs to show that  $F(m) \leq \liminf_{n \rightarrow \infty} F(m_n)$ . One can assume without loss of generality that  $F(m_n) < \infty$  for all  $n$ . By definition of  $\mathcal{T}_c(\cdot|\mu)$ , for all  $n \in \mathbb{N}^*$ , there exists  $\pi_n \in \Pi(\mu, m_n)$  such that  $I_c[\pi_n] - 1/n \leq \mathcal{T}_c(m_n|\mu) \leq I_c[\pi_n]$ . Since  $m_n$  is a converging sequence, the set  $\{m_n; n \in \mathbb{N}^*\} \cup \{m\}$  is relatively compact. Therefore, according to Theorem 8.3, for some arbitrary fixed point  $x_0 \in X$ , for all  $\varepsilon > 0$ , there exists a compact set  $K_\varepsilon \subset X$  such that

$$\sup_{n \in \mathbb{N}^*} \int_{X \setminus K_\varepsilon} 1 + \gamma(d(x_0, y)) m_n(dy) \leq \varepsilon \quad \text{and} \quad \int_{X \setminus K_\varepsilon} 1 + \gamma(d(x_0, x)) \mu(dx) \leq \varepsilon.$$

Therefore, letting  $M := \sup_{n \in \mathbb{N}^*} \int \gamma(d(x_0, x)) m_n(dx) < \infty$  and  $K_\varepsilon^c := X \setminus K_\varepsilon$ , it holds

$$\begin{aligned} & \int_{X \times X \setminus (K_\varepsilon \times K_\varepsilon)} 1 + \gamma(d(x_0, x)) + \gamma(d(x_0, y)) \pi_n(dxdy) \\ & \leq \int_{X \times K_\varepsilon^c} 1 + \gamma(d(x_0, x)) + \gamma(d(x_0, y)) \pi_n(dxdy) \\ & \quad + \int_{K_\varepsilon^c \times X} 1 + \gamma(d(x_0, x)) + \gamma(d(x_0, y)) \pi_n(dxdy) \\ & \leq m_n(K_\varepsilon^c) \int \gamma(d(x_0, x)) \mu(dx) + \int_{K_\varepsilon^c} 1 + \gamma(d(x_0, y)) m_n(dy) \\ & \quad + \int_{K_\varepsilon^c} 1 + \gamma(d(x_0, x)) \mu(dx) + \mu(K_\varepsilon^c)M \leq \varepsilon \left( 2 + M + \int \gamma(d(x_0, x)) \mu(dx) \right). \end{aligned}$$

So according to Theorem 8.3, it follows that  $\{\pi_n; n \in \mathbb{N}^*\}$  is relatively compact. Extracting a subsequence if necessary, one can assume without loss of generality that  $\pi_n$  converges to some  $\pi^* \in \mathcal{P}_\gamma(X \times X)$ . This  $\pi^*$  has the correct marginals  $\mu$  and  $m$ . Furthermore, denoting by  $\ell = \liminf_{n \rightarrow \infty} I_c[\pi_n] = \liminf_{n \rightarrow \infty} \mathcal{T}_c(m_n|\mu)$ , we see that, for all  $r > 0$ ,

$$\pi_n \in \{\pi \in \mathcal{P}_\gamma(X \times X); \pi(dx \times X) = \mu(dx) \text{ and } I_c[\pi] \leq \ell + r\} := A_{\ell+r},$$

for infinitely many  $n \in \mathbb{N}^*$ . By assumption  $(C_1)$ , the set  $A_{\ell+r}$  is closed for the topology  $\sigma(\mathcal{P}_\gamma(X \times X))$ . Therefore, the limit  $\pi^*$  also belongs to  $A_{\ell+r}$ . In other words,

$$F(m) = \mathcal{T}_c(m|\mu) \leq I_c[\pi^*] \leq \liminf_{n \rightarrow \infty} \mathcal{T}_c(m_n|\mu) + r, \quad \forall r > 0.$$

Since  $r > 0$  is arbitrary, this concludes the proof of the lower-semicontinuity of  $F$ .

According to Lemma 8.2 the topological dual space of  $\mathcal{M}_\gamma(X)$  can be identified with the set of linear functionals  $m \mapsto \int \varphi dm$ , where  $\varphi \in \Phi_\gamma(X)$ . Applying Theorem 8.1 together with Lemma 8.2 we conclude that

$$F(m) = \sup_{\varphi \in \Phi_\gamma(X)} \left\{ \int \varphi dm - F^*(\varphi) \right\} = \sup_{\varphi \in \Phi_\gamma(X)} \left\{ \int -\varphi dm - F^*(-\varphi) \right\}, \quad \forall m \in \mathcal{P}_\gamma(X).$$

Now we show that the last supremum can be restricted to  $\Phi_{\gamma,b}(X)$ . Observe that

$$\begin{aligned} F^*(-\varphi) &= \sup_{m \in \mathcal{P}_\gamma(X)} \left\{ \int -\varphi dm - F(m) \right\} \\ &= \sup_{k \in \mathbb{R}} \sup_{m \in \mathcal{P}_\gamma(X)} \left\{ \int -(\varphi \vee k) dm - F(m) \right\} = \sup_{k \in \mathbb{R}} F^*(-(\varphi \vee k)), \end{aligned}$$

so that for all  $\varphi \in \Phi_\gamma(X)$  and  $m \in \mathcal{P}_\gamma(X)$ , we have

$$\int -\varphi dm - F^*(-\varphi) = \lim_{k \rightarrow -\infty} \int -(\varphi \vee k) dm - F^*(-(\varphi \vee k)).$$

Therefore,

$$F(m) = \sup_{\varphi \in \Phi_\gamma(X)} \left\{ \int -\varphi dm - F^*(-\varphi) \right\} \leq \sup_{\varphi \in \Phi_{\gamma,b}(X)} \left\{ \int -\varphi dm - F^*(-\varphi) \right\},$$

and since the other inequality is obvious, the two quantities are equal. To conclude the proof it remains to show that

$$(8.4) \quad F^*(-\varphi) = - \int R_c \varphi(x) \mu(dx), \quad \forall \varphi \in \Phi_{\gamma,b}(X).$$

For all  $\varphi \in \Phi_{\gamma,b}$ , it holds

$$\begin{aligned} F^*(-\varphi) &= \sup_{m \in \mathcal{P}_\gamma(X)} \left\{ \int -\varphi dm - \mathcal{T}_c(m|\mu) \right\} = \sup_{m \in \mathcal{P}_\gamma(X)} \sup_{\pi \in \Pi(\mu,m)} \left\{ \int -\varphi dm - I_c[\pi] \right\} \\ &= \sup \left\{ \int \left[ \int -\varphi(y) p_x(dy) - c(x, p_x) \right] \mu(dx); \right. \\ &\quad \left. (p_x)_{x \in X} \text{ probability kernel such that } \mu p \in \mathcal{P}_\gamma(X) \right\} \\ &= - \inf \left\{ \int \left[ \int \varphi(y) p_x(dy) + c(x, p_x) \right] \mu(dx); \right. \\ &\quad \left. (p_x)_{x \in X} \text{ probability kernel such that } \mu p \in \mathcal{P}_\gamma(X) \right\}. \end{aligned}$$

By definition  $R_c \varphi(x) = \inf_{p \in \mathcal{P}_\gamma(X)} \{ \int \varphi dp + c(x, p) \}$ . Therefore, one has

$$F^*(-\varphi) \leq - \int R_c \varphi(x) \mu(dx).$$

Let us show the converse inequality. One can assume without loss of generality that  $\int R_c \varphi(x) \mu(dx) \in (-\infty, \infty)$ . For all  $\varepsilon > 0$  and  $x \in X$ , consider the set  $M_x^\varepsilon$  defined by

$$M_x^\varepsilon := \left\{ p \in \mathcal{P}_\gamma(X); \int \varphi dp + c(x, p) \leq R_c \varphi(x) + \varepsilon \right\}.$$

Note that since  $\varphi$  is bounded from below and  $c \geq 0$ ,  $R_c \varphi(x) > -\infty$  for all  $x \in X$  and so  $M_x^\varepsilon$  is non empty for all  $\varepsilon > 0$ .

Assume that for all  $\varepsilon > 0$ , there exists a measurable kernel  $X \rightarrow \mathcal{P}_\gamma(X) : x \mapsto p_x^\varepsilon$  such that for all  $x \in X$ ,  $p_x^\varepsilon \in M_x^\varepsilon$ . Then, if  $\varphi$  is bounded below by  $k$ , one sees that  $\int c(x, p_x^\varepsilon) \mu(dx) \leq -k + \varepsilon + \int R_c \varphi d\mu < \infty$ . According to condition  $(C_4)$  one concludes that  $\nu^\varepsilon = \mu p^\varepsilon \in \mathcal{P}_\gamma(X)$ . So it holds

$$F^*(-\varphi) \geq - \int \int \varphi(y) p_x^\varepsilon(dy) + c(x, p_x^\varepsilon) \mu(dx) \geq - \int R_c \varphi(x) \mu(dx) - \varepsilon,$$

which gives the desired inequality when  $\varepsilon \rightarrow 0$ .

When the condition  $(C_3)$  holds, the kernel  $p_x^\varepsilon$  is obtained by applying the elementary measurable selection result of Lemma 8.5 below. Indeed, note that the function  $H(x, p) = \int \varphi dp + c(x, p)$  is continuous (and thus upper-semicontinuous), and that  $Y = \mathcal{P}_\gamma(X)$  equipped with the topology  $\sigma(\mathcal{P}_\gamma(X))$  is metrizable (for instance, by the Kantorovich metric  $W_r$  if  $\gamma = \gamma_r$ , or the Lévy-Prokhorov distance for the usual weak-topology if  $\gamma = \gamma_0$ ) and separable (see [37, Theorem 6.18], [4, Proposition 7.20]).

Under condition  $(C'_3)$ , the space  $X$  is compact and the function  $H$  defined above is lower-semicontinuous. The selection Lemma 8.6 below ensures that there exists a measurable



kernel  $X \rightarrow \mathcal{P}_\gamma(X) : x \mapsto p_x$  such that  $R_c\varphi(x) = \inf_{p \in \mathcal{P}_\gamma(X)} H(x, p) = H(x, p_x)$ . The conclusion easily follows.

Under condition  $(C_3'')$ ,  $X$  is a countable set of isolated points. So all subsets of  $X$  are open (the topology on  $X$  is thus the discrete one) and all functions are measurable (and even continuous). Therefore by choosing for each  $x$  in  $X$ , some element  $p_x^\varepsilon$  in the non-empty set  $M_x^\varepsilon$ , we get a measurable kernel  $X \rightarrow \mathcal{P}_\gamma(X) : x \mapsto p_x^\varepsilon$ . The same conclusion follows.

To complete the proof, one needs to justify that Condition  $(C_1)$  follows from Condition  $(C_3'')$ . Assume that  $(X, d)$  is a countable set of isolated points and that for all  $x \in X$ , the function  $p \mapsto c(x, p)$  is lower-semicontinuous and let us show that  $\pi \mapsto I_c[\pi]$  is lower semicontinuous on  $\Pi(\mu, \cdot)$ . Let  $(\pi_n)_n$  be a sequence in  $\Pi(\mu, \cdot)$  converging to some  $\pi$  for the topology  $\sigma(\mathcal{P}_\gamma(X \times X))$ . Write  $\pi_n(dx dy) = p_{x,n}(dy) \mu(dx)$  and denote by  $\nu_n$  (resp.  $\nu$ ) the second marginal of  $\pi_n$  (resp.  $\pi$ ). The sequence  $\nu_n$  converges to  $\nu$ , therefore it is relatively compact and so according to Theorem 8.3, for all  $\varepsilon > 0$ , there is some compact  $K_\varepsilon \subset X$  (i.e. a finite set) such that  $\int_{K_\varepsilon^c} \gamma(d(x_o, y)) \nu_n(dy) \leq \varepsilon$ , where  $x_o$  is some fixed point in  $X$ . In other words,

$$\sum_{y \in K_\varepsilon^c} \sum_{x \in X} \gamma(d(x_o, y)) p_{x,n}(\{y\}) \mu(\{x\}) \leq \varepsilon$$

In particular, for all  $x \in X$  in the support of  $\mu$ , it holds  $\sum_{y \in K_\varepsilon^c} \gamma(d(x_o, y)) p_{x,n}(\{y\}) \leq \varepsilon / \mu(\{x\})$ , and so, according to Theorem 8.3,  $\{p_{x,n}; n \in \mathbb{N}\}$  is relatively compact. Without loss of generality (extracting a subsequence if necessary), one can assume that  $I_c[\pi_n] = \int c(x, p_{x,n}) \mu(dx)$  converges. Since for all  $x$  in the support of  $\mu$ ,  $\{p_{x,n}; n \in \mathbb{N}\}$  is relatively compact, the classical diagonal extraction argument enables us to construct an increasing map  $\sigma : \mathbb{N} \rightarrow \mathbb{N}$  such that  $\tilde{p}_{x, \sigma(n)}$  converges to some  $p_x \in \mathcal{P}_\gamma(X)$  as  $n \rightarrow \infty$ , for all  $x$  in the support of  $\mu$ . Finally, using Fatou's lemma and the lower-semicontinuity of  $p \mapsto c(x, p)$ , one gets

$$\lim_{n \rightarrow \infty} I_c[\pi_n] = \lim_{n \rightarrow \infty} \int c(x, p_{x, \sigma(n)}) \mu(dx) \geq \int \liminf_{n \rightarrow \infty} c(x, p_{x, \sigma(n)}) \mu(dx) \geq \int c(x, p_x) \mu(dx).$$

It remains to show that the last term is equal to  $I_c[\pi]$ . But if  $f : X \times X \rightarrow \mathbb{R}$  is bounded (continuous), then by dominated convergence

$$\begin{aligned} \int f(x, y) \pi(dx dy) &= \lim_{n \rightarrow \infty} \int f(x, y) \pi_{\sigma(n)}(dx dy) = \lim_{n \rightarrow \infty} \int \left( \int f(x, y) p_{x, \sigma(n)}(dy) \right) \mu(dx) \\ &= \int \left( \int f(x, y) p_x(dy) \right) \mu(dx). \end{aligned}$$

Since this holds for all  $f$ , one concludes that  $p_x(dy) \mu(dx) = \pi(dx dy)$  and so in particular,  $\int c(x, p_x) \mu(dx) = I_c[\pi]$  which completes the proof.  $\square$

In the proof of Theorem 3.5 we used the following results, elementary proofs of which can be found in [4] (see Proposition 7.34 and Proposition 7.33).

**Lemma 8.5.** *Let  $X$  be a metrizable space,  $Y$  a metrizable and separable space and  $H : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$  be an upper-semicontinuous function. Denoting by  $\overline{H}(x) = \inf_{y \in Y} H(x, y) \in \mathbb{R} \cup \{\pm\infty\}$ , for all  $\varepsilon > 0$ , there exists a measurable function  $x \mapsto s^\varepsilon(x)$  such that*

$$H(x, s^\varepsilon(x)) \leq \begin{cases} \overline{H}(x) + \varepsilon & \text{if } \overline{H}(x) > -\infty \\ -1/\varepsilon & \text{if } \overline{H}(x) = -\infty. \end{cases}$$

**Lemma 8.6.** *Let  $X$  be a metrizable space,  $Y$  a compact metrizable space and  $H : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$  be a lower-semicontinuous function. Then there exists a measurable function*

$x \mapsto s(x)$  such that for all  $x \in X$

$$H(x, s(x)) = \inf_{y \in Y} H(x, y).$$

### 8.3. Proofs of Corollaries 3.7, 3.9, 3.11 and 3.12.

#### 8.3.1. Proof of Corollary 3.7.

*Proof of Corollary 3.7.* First assume that  $\omega: X \times X \rightarrow [0, \infty)$  is continuous and bounded from above. Then  $c(x, p) = \int \omega(x, y) p(dy)$  is convex in  $p$  and continuous on  $X \times \mathcal{P}(X)$ , with  $\mathcal{P}(X)$  equipped with the usual weak topology. Moreover  $I_c[\pi] = \int \omega(x, y) \pi(dxdy)$  and so  $\pi \mapsto I_c[\pi]$  is continuous on  $\mathcal{P}(X \times X)$ . So assumptions  $(C_1), (C_2), (C_3), (C_4)$  of Theorem 3.5 are fulfilled with  $\mathcal{P}_\gamma(X) = \mathcal{P}(X)$  and  $\Phi_{\gamma, b} = \Phi_0$ . It follows that

$$\mathcal{T}_\omega(\nu, \mu) = \sup_{\varphi \in \Phi_0(X)} \left\{ \int R_c \varphi(x) \mu(dx) - \int \varphi(y) \nu(dy) \right\},$$

with

$$R_c \varphi(x) = \inf_{p \in \mathcal{P}(X)} \left\{ \int \varphi(y) + \omega(x, y) p(dy) \right\} = \inf_{y \in X} \{ \varphi(y) + \omega(x, y) \} = Q_c \varphi(x),$$

which completes the proof in the case of a bounded continuous cost function. Once Kantorovich duality is established for bounded continuous cost functions, one can apply a rather standard approximation argument to extend the duality to lower-semicontinuous cost functions. This is explained for instance in [36, Point 3 in the proof of Theorem 1.3].  $\square$

#### 8.3.2. Proof of Corollary 3.9.

*Proof of Corollary 3.9.* Depending on the assumption on the space and on  $\alpha$ , one needs to verify that Condition  $(C)$ ,  $(C')$  or  $(C'')$  of Theorem 3.5 is satisfied. We distinguish between the different cases.

If  $\alpha: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is convex and continuous, the cost  $c(x, p) = \alpha(\int \gamma(d(x, y)) p(dy))$  is clearly convex with respect to  $p$  and, by definition of the topology  $\sigma(\mathcal{P}_\gamma(X))$ , it is continuous on  $X \times \mathcal{P}_\gamma(X)$  (equipped with the product topology). So assumptions  $(C_2), (C_3)$  of Theorem 3.5 are fulfilled. Condition  $(C_4)$  follows at once from Jensen's inequality. As for Condition  $(C_1)$ , let us set  $\alpha(t) = +\infty$  for  $t < 0$  so that  $\alpha$  is lower-semicontinuous on  $\mathbb{R}$ . According to the Fenchel-Legendre duality Theorem 8.1,

$$\alpha(t) = \sup_{s \geq \alpha'(0)} \{st - \alpha^*(s)\} = \sup_{s \geq 0} \{st - \alpha^*(s)\},$$

where  $\alpha'(0)$  is the non-negative right-derivative of  $\alpha$  at point 0, and  $\alpha^*(s) = \sup_{t \geq 0} \{st - \alpha(t)\}$ . So

$$\begin{aligned} c(x, p) &= \sup_{s \geq 0} \int s\gamma(d(x, y)) - \alpha^*(s) p(dy) = \sup_{(s, t) \in \text{epi}(\alpha^*)} \int s\gamma(d(x, y)) - t p(dy) \\ &= \sup_{k \in \mathbb{N}} \int \varphi_k(x, y) p(dy), \end{aligned}$$

with  $\varphi_0 = 0$  and  $\varphi_k(x, y) = s_k \gamma(d(x, y)) - t_k$ ,  $k \geq 1$  where  $(s_k, t_k)_{k \geq 1}$  is any dense subset of  $\text{epi}(\alpha^*) = \{(s, t) \in [0, \infty) \times \mathbb{R}; t \geq \alpha^*(s)\}$ . For all  $k \in \mathbb{N}$ ,  $\varphi_k \in \Phi_\gamma(X \times X)$  and so according to Proposition 3.3, the cost function  $c$  verifies  $(C_1)$ .

If  $\alpha: \mathbb{R} \rightarrow [0, +\infty]$  is convex and lower-semicontinuous, then  $c$  is also clearly convex with respect to  $p$  (hence Condition  $(C_2)$  is satisfied). Since  $\gamma$  is lower-semicontinuous, there exists an increasing sequence  $(\gamma_N)_{N \in \mathbb{N}}$  of Lipschitz continuous functions  $\gamma_N: \mathbb{R}_+ \rightarrow \mathbb{R}_+$

that converges to  $\gamma$  (for example  $\gamma_N(u) = \inf_{v \in \mathbb{R}} \{\gamma(v) + N|u - v|\}$ ). By using the Fenchel-Legendre duality for  $\alpha$  as above and by monotone convergence, one has

$$c(x, p) = \sup_{(s, t) \in \text{epi}(\alpha^*)} \sup_{N \in \mathbb{N}} \int s \gamma_N(d(x, y)) - t p(dy) = \sup_{k \in \mathbb{N}} \int \varphi_k(x, y) p(dy),$$

with  $\varphi_0 = 0$  and  $\varphi_k(x, y) = s_{\ell(k)} \gamma_{N(k)}(d(x, y)) - t_{\ell(k)}$ ,  $k \geq 1$  where  $(s_l, t_l)_{l \in \mathbb{N}}$  is any dense subset of  $\text{epi}(\alpha^*) = \{(s, t) \in [0, \infty) \times \mathbb{R}; t \geq \alpha^*(s)\}$ , and the map  $\mathbb{N}^* \ni k \mapsto (N(k), \ell(k)) \in \mathbb{N} \times \mathbb{N}$  is one to one. By Proposition 3.3, the conditions  $(C_1)$ , and  $(C'_3)$  are fulfilled when  $X$  is compact, and respectively  $(C''_3)$  when  $X$  is a countable set of isolated points. Condition  $(C_4)$  is again a consequence of Jensen's inequality.

The result of the corollary is finally a direct consequence of Theorem 3.5.  $\square$

### 8.3.3. Proof of Corollary 3.11.

*Proof of Corollary 3.11.*

(1) The proof of the first point is similar to that of Corollary 3.9. Namely, if  $\theta: \mathbb{R}^m \rightarrow \mathbb{R}$  is a convex function, assumptions  $(C_2)$ ,  $(C_3)$  are satisfied with  $\gamma = \gamma_1$ . Since  $\theta(x) \geq a\|x\| + b$  for some  $a > 0$  and  $b \in \mathbb{R}$ , condition  $(C_4)$  follows easily from Jensen's inequality. Finally, using Fenchel-Legendre duality for  $\theta$ , one sees that

$$c(x, p) = \theta \left( x - \int y p(dy) \right) = \sup_{(s, t) \in \text{epi}(\theta^*)} \int s \cdot (x - y) - t p(dy),$$

with  $\text{epi}(\theta^*) = \{(s, t) \in \mathbb{R}^m \times \mathbb{R}; \theta^*(s) \leq t\}$ . Taking a dense countable subset  $(s_k, t_k)_{k \geq 1}$  of  $\text{epi}(\theta^*)$ , one concludes that  $c(x, p) = \sup_{k \in \mathbb{N}} \int \varphi_k(x, y) p(dy)$ , with  $\varphi_0 = 0$  and  $\varphi_k(x, y) = s_k(x - y) - t_k$ . These functions belong to  $\Phi_1(X \times X)$ , so according to Proposition 3.3, the cost function  $c$  verifies  $(C_1)$ .

If  $\theta: \mathbb{R}^m \rightarrow (-\infty, +\infty]$  is a lower-semicontinuous convex function, we show similarly that  $(C_1)$ ,  $(C_2)$ ,  $(C_4)$  are fulfilled, along with  $(C'_3)$  when  $X$  is compact, and respectively  $(C''_3)$  when  $X$  is discrete.

(2) Let  $\varphi \in \Phi_{1,b}(\mathbb{R}^m)$ , it holds for all  $x \in \mathbb{R}^m$ ,

$$\overline{Q}_\theta \varphi(x) = \inf_{p \in \mathcal{P}_1(\mathbb{R}^m)} \left\{ \int \varphi dp + \theta \left( x - \int y p(dy) \right) \right\} = \inf_{z \in \mathbb{R}^m} \{g(z) + \theta(x - z)\},$$

where

$$g(z) := \inf \left\{ \int \varphi dp; p \in \mathcal{P}_1(\mathbb{R}^m), \int y p(dy) = z \right\}, \quad z \in \mathbb{R}^m.$$

The function  $g$  is easily seen to be convex on  $\mathbb{R}^m$ . This implies that  $g \leq \overline{\varphi}$ . Let us show that  $g \geq \overline{\varphi}$ . Since  $\varphi$  is bounded from below there is some  $a \in \mathbb{R}$  such that  $\varphi(y) \geq a$ , for all  $y \in \mathbb{R}^m$ . Then by definition of  $\overline{\varphi}$ , it holds  $\overline{\varphi}(y) \geq a$ , for all  $y \in \mathbb{R}^m$ . Since  $\overline{\varphi} \leq \varphi$ , it follows that  $\overline{\varphi}$  is finite everywhere. As a consequence, one can apply Jensen's inequality: if  $p \in \mathcal{P}_1(\mathbb{R}^m)$  is such that  $\int y p(dy) = z$ , then

$$\int \varphi(y) p(dy) \geq \int \overline{\varphi}(y) p(dy) \geq \overline{\varphi} \left( \int y p(dy) \right) = \overline{\varphi}(z).$$

Optimizing over  $p$ , one concludes that  $g(z) \geq \overline{\varphi}(z)$ , for all  $z \in \mathbb{R}^m$  and so finally  $g = \overline{\varphi}$ .

(3) Let  $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^m)$  and  $\varphi \in \Phi_{1,b}(\mathbb{R}^m)$ . According to Point (2), since  $\overline{\varphi} \leq \varphi$ , it holds

$$\int \overline{Q}_\theta \varphi d\mu - \int \varphi d\nu = \int Q_\theta \overline{\varphi} d\mu - \int \varphi d\nu \leq \int Q_\theta \overline{\varphi} d\mu - \int \overline{\varphi} d\nu.$$

The function  $\bar{\varphi}$  is convex, bounded from below and, since  $\varphi \in \Phi_1(\mathbb{R}^m)$ , satisfies  $\bar{\varphi}(x) \leq a + b\|x\|$ ,  $x \in \mathbb{R}^m$ , for some  $a, b \geq 0$ . This shows that  $\bar{\varphi} \in \Phi_{1,b}(\mathbb{R}^m)$ . From these considerations, it follows that

$$\begin{aligned} \bar{\mathcal{T}}_\theta(\nu|\mu) &\leq \sup \left\{ \int Q_\theta \bar{\varphi} d\mu - \int \bar{\varphi} d\nu; \varphi \in \Phi_{1,b}(\mathbb{R}^m) \right\} \\ &\leq \sup \left\{ \int Q_\theta \psi d\mu - \int \psi d\nu; \psi \in \Phi_{1,b}(\mathbb{R}^m) \text{ convex} \right\} \\ &\leq \sup \left\{ \int \bar{Q}_\theta \psi d\mu - \int \psi d\nu; \psi \in \Phi_{1,b}(\mathbb{R}^m) \right\} \\ &= \bar{\mathcal{T}}_\theta(\nu|\mu). \end{aligned}$$

The third inequality is a consequence of Point (2), since  $\psi = \bar{\psi}$  for all convex functions  $\psi \in \Phi_{1,b}(\mathbb{R}^m)$ . Remarking that a convex function belongs to  $\Phi_1(\mathbb{R}^m)$  if and only if it is Lipschitz, the proof of Point (3) is complete.

(4) We already know from Point (3) that for all  $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^m)$  it holds

$$\bar{\mathcal{T}}_1(\nu|\mu) = \sup \left\{ \int Q_1 \varphi d\mu - \int \varphi d\nu; \varphi \text{ convex Lipschitz bounded from below} \right\},$$

with  $Q_1 \varphi(x) = \inf_{y \in \mathbb{R}^m} \{\varphi(y) + \|x - y\|\}$ ,  $x \in \mathbb{R}^m$ . It is easy to check that if  $\varphi : \mathbb{R}^m \rightarrow \mathbb{R}$  is convex and bounded from below, so is  $Q_1 \varphi : \mathbb{R}^m \rightarrow \mathbb{R}$ . Being an infimum of 1-Lipschitz functions,  $Q_1 \varphi$  is itself 1-Lipschitz. Moreover, if  $\psi : \mathbb{R}^m \rightarrow \mathbb{R}$  is some 1-Lipschitz convex function, then  $Q_1 \psi = \psi$ ; namely, for all  $x \in \mathbb{R}^m$ , one has

$$0 \geq Q_1 \psi(x) - \psi(x) \geq \inf_{y \in \mathbb{R}^m} \{\psi(y) - \psi(x) + \|x - y\|\} \geq 0.$$

From these considerations, we conclude that

$$\begin{aligned} \bar{\mathcal{T}}_1(\nu|\mu) &= \sup \left\{ \int Q_1 \varphi d\mu - \int \varphi d\nu; \varphi \text{ convex Lipschitz bounded below} \right\} \\ &\leq \sup \left\{ \int \psi d\mu - \int \psi d\nu; \psi \text{ convex 1-Lipschitz bounded below} \right\} \\ &= \sup \left\{ \int Q_1 \psi d\mu - \int \psi d\nu; \psi \text{ convex 1-Lipschitz bounded below} \right\} \\ &\leq \sup \left\{ \int Q_1 \varphi d\mu - \int \varphi d\nu; \varphi \text{ convex Lipschitz bounded below} \right\}. \end{aligned}$$

This concludes the proof of Point (4).  $\square$

**8.3.4. Proof of Corollary 3.12.** We start with an alternative representation of  $c(x, p)$  that will be useful later on. We recall that  $c : X \times \mathcal{P}_\gamma(X) \rightarrow \mathbb{R}_+$  is defined by

$$c(x, p) = \int \beta \left( \gamma(d(x, y)) \frac{dp}{d\mu_0}(y) \right) \mu_0(dy)$$

if  $p \ll \mu_0$  on  $X \setminus \{x\}$  and  $+\infty$  otherwise, where  $\mu_0$  is a reference probability measure and  $\beta : \mathbb{R}_+ \rightarrow [0, \infty]$  is a lower-semicontinuous convex function such that  $\beta(0) = 0$  and  $\beta(x)/x \rightarrow \infty$  as  $x \rightarrow \infty$ . As before  $\gamma : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is a lower-semicontinuous function satisfying (2.1).

**Lemma 8.7.** *Let  $X$  be a metric space being either compact or a countable set of isolated points. The cost function  $c$  defined above satisfies the following duality identity:*

$$c(x, p) = \sup_{h \in \Phi_0(X), h \geq 0} \left\{ \int h(y) \gamma(d(x, y)) p(dy) - \int \beta^*(h)(y) \mu_0(dy) \right\},$$

where  $\beta^*$  denotes the Fenchel-Legendre transform of  $\beta$  defined by  $\beta^*(y) = \sup_{x \geq 0} \{xy - \beta(x)\}$ , for all  $y \in \mathbb{R}$ .

*Proof.* The proof is easily adapted from Theorem B.2 in [20].  $\square$

*Proof of Corollary 3.12.* First, we observe that Condition  $(C_2)$  is a simple consequence of the convexity of  $\beta$  and Condition  $(C_4)$  of Jensen's inequality. According to Lemma 8.7, it holds

$$(8.8) \quad \begin{aligned} c(x, p) &= \sup_{h \in \Phi_0(X), h \geq 0} \left\{ \int h(y) \gamma(d(x, y)) p(dy) - \int \beta^*(h)(y) \mu_0(dy) \right\} \\ &= \sup_{h \in \Phi_0(X), h \geq 0} \sup_{N \in \mathbb{N}} \int (h(y) \gamma_N(d(x, y)) - B^*(h)) p(dy), \end{aligned}$$

where  $(\gamma_N)_{N \in \mathbb{N}}$  is (as in the proof of Corollary 3.9) an increasing sequence of Lipschitz continuous functions converging to  $\gamma$  and  $B^*(h) = \int \beta^*(h) d\mu_0$ .

For all  $h \geq 0$ ,  $h \in \Phi_0(X)$ , and  $N \in \mathbb{N}$ , the function  $(x, y) \mapsto h(y) \gamma_N(d(x, y))$  is continuous. Therefore, the function  $p \mapsto \int h(y) \gamma(d(x, y)) p(dy)$  is continuous on  $X \times \mathcal{P}_\gamma(X)$ . Being a supremum of continuous functions,  $c$  is lower-semicontinuous on  $X \times \mathcal{P}_\gamma(X)$ . In particular, this shows  $(C'_3)$  and  $(C''_3)$ .

Next we will check that Condition  $(C_1)$  holds (in the compact case).

Since  $(X, d)$  is compact, the space  $\Phi_0(X)$  of continuous functions (equipped with the norm  $\|\cdot\|_\infty$ ) on  $X$  is separable (see [4, Proposition 7.7]). Let  $\{h_\ell, \ell \in \mathbb{N}\}$  be a countable dense subset of  $\Phi_0(X)$ . Since  $\beta^*$  is convex and finite on  $\mathbb{R}$  it is continuous on  $\mathbb{R}$ . Therefore, the function  $\Phi_0(X) \rightarrow \mathbb{R} : h \mapsto B^*(h)$  is continuous. It follows that

$$(8.9) \quad c(x, p) = \sup_{k \in \mathbb{N}} \int \varphi_k(x, y) p(dy), \quad \forall x \in X, \quad p \in \Phi_\gamma(X)$$

where  $\varphi_0 = 0$  and  $\varphi_k(x, y) = h_{\ell(k)}(y) \gamma_{N(k)}(d(x, y)) - B^*(h_{\ell(k)})$ ,  $k \geq 1$ , and  $\mathbb{N}^* \ni k \mapsto (\ell(k), N(k)) \in \mathbb{N} \times \mathbb{N}$  is one-to-one. Since, for all  $k \in \mathbb{N}$ , the function  $\varphi_k$  belongs to  $\Phi_\gamma(X, X)$ , the lower-semicontinuity of  $I_c$  follows from Proposition 3.3.

Corollary 3.12 now follows from Theorem 3.5.  $\square$

**8.4. Proof of Proposition 3.3.** The proof of Proposition 3.3 is adapted from [2, Theorem 2.34].

*Proof of Proposition 3.3.* The function  $p \mapsto c(x, p)$  is convex as a supremum of linear functions.

For all  $n \in \mathbb{N}$ , define  $c_n(x, p) := \sup_{k \leq n} \int \varphi_k(x, y) p(dy)$ . When  $n$  goes to  $\infty$ ,  $c_n(x, p)$  is a nondecreasing sequence converging to  $c$ . Let  $\pi \in \Pi(\mu, \cdot)$ ,  $\pi(dxdy) = p_x(dy) \mu(dx)$  such that (3.4) holds for  $\mu$ -almost all  $x$ . Defining  $I_{c_n}[\pi] = \int c_n(x, p_x) \mu(dx)$ , the monotone convergence theorem shows that  $I_c[\pi] = \sup_{n \in \mathbb{N}} I_{c_n}[\pi]$ . Since a supremum of lower-semicontinuous functions is itself lower-semicontinuous, it is enough to prove that  $I_{c_n}$  is lower-semicontinuous at point  $\pi$ . We will now prove such a property.

For  $\mu$ -almost all  $x$ , define  $\psi_k(x) = \int \varphi_k(x, y) p_x(dy)$ ,  $k \leq n$ . Then it holds

$$I_{c_n}[\pi] = \int \sup_{k \leq n} \psi_k(x) \mu(dx) = \sup_{(f_k)_{k \leq n}} \int \sum_{k=0}^n f_k(x) \psi_k(x) \mu(dx),$$

where the supremum runs over the set of continuous functions  $f_k$  taking values in  $[0, 1]$  and such that  $f_0 + \dots + f_n \leq 1$ . Let us admit this claim for a moment and finish the proof of the proposition. For all  $f_0, \dots, f_n$  as above, it holds

$$\int \sum_{k=0}^n f_k(x) \psi_k(x) \mu(dx) = \int \sum_{k=0}^n f_k(x) \varphi_k(x, y) \pi(dxdy).$$

Since  $\sum_{k=0}^n f_k \varphi_k \in \Phi_\gamma(X \times X)$ , the function  $\pi \mapsto \int \sum_{k=0}^n f_k \varphi_k d\pi$  is continuous on  $\Pi(\mu, \cdot)$ . Since a supremum of continuous functions is lower-semicontinuous, this proves that  $I_{c_n}$  is lower-semicontinuous at point  $\pi$ .

It remains to prove the claim. Obviously, if  $f_0, f_1, \dots, f_n$  take values in  $[0, 1]$  and are such that  $\sum_{k=0}^n f_k \leq 1$ , then it holds

$$\begin{aligned} \int \sum_{k=0}^n f_k(x) \psi_k(x) \mu(dx) &\leq \int \sum_{k=0}^n f_k(x) [\psi_k]_+(x) \mu(dx) \leq \int \sup_j [\psi_j]_+(x) \sum_{k=0}^n f_k(x) \mu(dx) \\ &\leq \int \sup_j [\psi_j]_+(x) \mu(dx) = I_{c_n}[\pi], \end{aligned}$$

where the last equality comes from the fact that  $\sup_j [\psi_j]_+ = \sup_j \psi_j$  since  $\varphi_0 = 0$  and  $\psi_0 = 0$ . This shows that  $I_{c_n}[\pi] \geq \sup_{(f_k)_{k \leq n}} \int \sum_{k=0}^n f_k(x) \psi_k(x) \mu(dx)$ .

To prove the converse inequality, let  $A_k = \{x \in X; [\psi_k]_+ = \sup_j [\psi_j]_+(x)\}$  for all  $k \leq n$ , and define recursively  $B_0 = A_0$ ,  $B_k = A_k \setminus (B_0 \cup \dots \cup B_{k-1})$ . Then it holds

$$I_{c_n}[\pi] = \sum_{k=0}^n \int_{B_k} [\psi_k]_+(x) \mu(dx).$$

When  $(X, d)$  is a discrete space, the functions  $f_k = \mathbf{1}_{B_k}$  are continuous and  $\sum_{k=0}^n f_k = 1$ . Since  $\psi_k$  is non-negative on  $A_k$ , one has

$$I_{c_n}[\pi] = \sum_{k=0}^n \int f_k(x) \psi_k(x) \mu(dx),$$

and the claim follows in this case.

Assume now that  $(X, d)$  is complete and separable. For all  $k \leq n$ , consider the finite Borel measure  $\mu_k(dx) = [\psi_k]_+(x) \mu(dx)$ . Let  $\varepsilon > 0$ ; since finite Borel measures on a complete separable metric space are inner regular (see for instance [27, Theorems 3.1 and 3.2]), for all  $k \leq n$  there is a compact set  $C_k \subset B_k$  such that  $\mu_k(B_k) \leq \mu_k(C_k) + \varepsilon/(n+1)$ . So it holds

$$I_{c_n}[\pi] = \sum_{k=0}^n \int_{B_k} [\psi_k]_+(x) \mu(dx) \leq \sum_{k=0}^n \int_{C_k} [\psi_k]_+(x) \mu(dx) + \varepsilon = \sum_{k=0}^n \int_{C_k} \psi_k(x) \mu(dx) + \varepsilon.$$

The compact sets  $C_k$  are pairwise disjoint, so  $\delta_o = \min_{i \neq j} d(C_i, C_j) > 0$ . Consider the family of continuous functions  $f_{k,\delta} : X \rightarrow [0, 1]$  defined by

$$f_{k,\delta}(x) = \left[ 1 - \frac{d(x, C_k)}{\delta} \right]_+, \quad x \in X, \quad k \leq n, \quad \delta > 0.$$

When  $\delta < \delta_o/2$ , for any  $x \in X$ , at most one of the functions is not zero at  $x$  and therefore  $\sum_{k=0}^n f_{k,\delta}(x) \leq 1$ . Passing to the limit when  $\delta \rightarrow 0$ , we see that

$$\sum_{k=0}^n \int f_{k,\delta}(x) \psi_k(x) \mu(dx) \rightarrow \sum_{k=0}^n \int_{C_k} \psi_k(x) \mu(dx).$$

So if  $\delta$  is small enough it holds

$$I_{c_n}[\pi] \leq \sum_{k=0}^n \int f_{k,\delta}(x) \psi_k(x) \mu(dx) + 2\varepsilon.$$

Taking the supremum over all possible functions  $f_k$ , and then letting  $\varepsilon$  go to 0, gives the desired inequality  $I_{c_n}[\pi] \leq \sup_{(f_k)_{k \leq n}} \int \sum_{k=0}^n f_k(x) \psi_k(x) \mu(dx)$ , and completes the proof.  $\square$

## APPENDIX A.

In this appendix we prove some technical results on the function  $\theta_{\rho,t}$ .

Recall that, given  $\rho \in (0, 1)$ ,

$$u_{\rho}(x) = \begin{cases} \frac{1-\rho(1-x)}{\rho} \log \frac{1-\rho(1-x)}{1-\rho} + (1-x) \log(1-x) & \text{if } -\frac{1-\rho}{\rho} \leq x \leq 1 \\ +\infty & \text{otherwise} \end{cases}$$

and given  $t \in (0, 1)$ , define  $\theta_{\rho,t}: \mathbb{R} \rightarrow \mathbb{R}$  as

$$\theta_{\rho,t}(h) = \frac{1}{t(1-t)} \inf_{\tau \geq 1} \begin{cases} \frac{1}{\tau} [(1-t)u_{\rho}(1-\tau) + tu_{\rho}(1-\tau(1-h))] & \text{if } h \geq 0 \\ \frac{1}{\tau} [(1-t)u_{1-\rho}(1-\tau) + tu_{1-\rho}(1-\tau(1+h))] & \text{if } h < 0. \end{cases}$$

Observe that

$$(A.1) \quad \theta_{\rho,t}(h) = \frac{1}{t(1-t)} \begin{cases} \inf_{1 \leq x \leq \frac{1}{\rho}} \psi_{t,\rho}(h, x) & \text{if } h \in [0, 1] \\ \inf_{1 \leq x \leq \frac{1}{1-\rho}} \psi_{t,1-\rho}(-h, x) & \text{if } h \in [-1, 0] \\ +\infty & \text{otherwise} \end{cases}$$

where

$$\psi_{t,\rho}(h, x) := \frac{1}{x} [(1-t)u_{\rho}(1-x) + tu_{\rho}(1-x(1-h))].$$

Define  $\theta_{\rho,t}$  for  $t = 0, 1$  as the point-wise limit of  $\theta_{\rho,t}$  as  $t$  tends to  $0, 1$ . That is, (see [30, Proposition 2.4]),

$$\theta_{\rho,0}(h) = \begin{cases} u_{\rho}(h) & \text{if } h \geq 0 \\ u_{1-\rho}(-h) & \text{if } h < 0 \end{cases}$$

and

$$\theta_{\rho,1}(h) = \begin{cases} \frac{1}{\rho} \left[ (1-\rho-h) \log \frac{1-\rho-h}{1-\rho} - (1-h) \log(1-h) \right] & \text{if } h \in [0, 1-\rho] \\ \frac{1}{1-\rho} \left[ (\rho+h) \log \frac{\rho+h}{\rho} - (1+h) \log(1+h) \right] & \text{if } h \in [-\rho, 0] \\ +\infty & \text{otherwise.} \end{cases}$$

**Lemma A.2.** *for all  $t \in [0, 1]$  the mapping  $\mathbb{R} \ni h \mapsto \theta_{\rho,t}(h)$  is convex and compares to  $h^2$  on  $[-1, 1]$ .*

*Proof.* We will first prove that  $\theta_{\rho,t}$  is convex. The limiting cases  $t \rightarrow 0, 1$  can be deduced from the general case  $t \in (0, 1)$  on which we focus now. We may start with  $h \in [0, 1]$ . For simplicity, and when there is no confusion, we may often drop the indices  $\rho$  and  $t$  in the above notations, setting  $\psi := \psi_{t,\rho}$ , etc.

A simple computation leads to

$$\frac{\partial}{\partial x} \psi(h, x) = -\frac{1}{\rho x^2} \left( (1-t) \log \frac{1-\rho x}{1-\rho} + t \log \frac{1-\rho x(1-h)}{1-\rho} \right) =: -\frac{1}{\rho x^2} H(x)$$

and

$$H'(x) = -\frac{\rho}{(1-\rho x)(1-\rho x(1-h))} (1-th - \rho x(1-h)),$$

for any  $x \in [1, 1/\rho]$ . Since  $h \in [0, 1]$ , the mapping  $G: [1, 1/\rho] \ni x \mapsto 1-th - \rho x(1-h)$  is decreasing and so  $G(x) \geq G(1/\rho) = h(1-t) > 0$  for any  $x \in [1, 1/\rho]$ . It follows that  $H' < 0$  on  $[1, 1/\rho]$  and therefore that  $H$  is decreasing. Now  $H(1) = t \log \frac{1-\rho(1-h)}{1-\rho} \geq 0$  (since  $h \in [0, 1]$ ) and  $\lim_{x \rightarrow 1/\rho} H(x) = -\infty$ , so that there exists a unique point  $\bar{x} = \bar{x}_{\rho,t}(h) \in [1, 1/\rho]$  such that  $H(\bar{x}) = \frac{\partial}{\partial x} \psi(h, \bar{x}) = 0$ , and (since  $H$  and  $x \mapsto \frac{\partial}{\partial x} \psi(h, x)$  have opposite signs)

$$\theta_{\rho,t}(h) = \frac{1}{t(1-t)} \psi_{t,\rho}(h, \bar{x}_{\rho,t}(h)), \quad h \in [0, 1].$$

Moreover,  $\bar{x}$  is unequivocally (and implicitly) defined by the equation

$$(A.3) \quad (1 - \rho\bar{x})^{1-t}(1 - \rho\bar{x}(1 - h))^t = 1 - \rho.$$

Now, since  $\frac{\partial}{\partial x}\psi(h, \bar{x}) = 0$ , we get after simple computations<sup>6</sup>

$$\begin{aligned} t(1-t)\theta''(h) &= \frac{\partial^2}{\partial h^2}\psi(h, \bar{x}) + \frac{\partial^2}{\partial x\partial h}\psi(h, \bar{x}) \cdot \bar{x}'(h) \\ &= \frac{t}{(1-h)(1-\rho\bar{x}(1-h))} + \frac{-t}{\bar{x}(1-\rho\bar{x}(1-h))} \cdot \bar{x}'(h) \\ &= \frac{t[\bar{x} - (1-h)\bar{x}'(h)]}{\bar{x}(1-h)(1-\rho\bar{x}(1-h))}. \end{aligned}$$

It follows by differentiating Equation (A.3) that

$$(A.4) \quad \bar{x}'(h) = \frac{t\bar{x}(1-\rho\bar{x})}{1-th-\rho\bar{x}(1-h)},$$

which in turn implies (after some algebra) that

$$\theta''(h) = \frac{1}{(1-h)(1-th-\rho\bar{x}(1-h))} \geq 0,$$

since  $h \in [0, 1)$ ,  $t \in (0, 1)$  and  $1-th-\rho\bar{x}(1-h) \geq 1-th-(1-h) = h(1-t) \geq 0$  (noting that  $\rho\bar{x} \leq 1$ ).

By construction of  $\theta_{\rho,t}$  we also have  $\theta''_{\rho,t} \geq 0$  on  $(-1, 0]$ . Hence, all what remains to prove is that  $\theta_{\rho,t}$  is continuous at  $h = 0$  and  $\lim_{h \rightarrow 0^-} \theta'_{\rho,t}(h) \leq \lim_{h \rightarrow 0^+} \theta'_{\rho,t}(h)$ . By the above computations, we deduce that  $\bar{x}(0^+) = \bar{x}(0^-) = 1$ , so that  $\theta_{\rho,t}(0^-) = \frac{u_{1-\rho}(0)}{t(1-t)} = 0 = \frac{u_{\rho}(0)}{t(1-t)} = \theta_{\rho,t}(0^+)$  (since  $u_{\rho}(0) = 0$ ). Furthermore, since  $\frac{1}{t}\frac{\partial}{\partial h}\psi_{\rho,t}(h, x) = \log(1-\rho x(1-h)) - \log[x(1-\rho)(1-h)]$ , we have  $\lim_{h \rightarrow 0^-} \theta'_{\rho,t}(h) = \lim_{h \rightarrow 0^+} \theta'_{\rho,t}(h) = 0$ . This ends the proof of the convexity of  $\theta_{\rho,t}$ .

That  $\theta_{\rho,t}$  compares to  $h^2$  on  $[-1, 1]$  is a simple consequence of the fact that,  $(1-\rho)\theta''_{\rho,t}(h) = 1 + o(1)$ , when  $h \rightarrow 0^+$  (and similarly, but with a different multiplicative factor, when  $h \rightarrow 0^-$ ).  $\square$

*Proof of Claim (7.9) and Claim (7.19).* We start with the proof of (7.9),

$$\lim_{n \rightarrow \infty} n\theta_{\rho_n,t} \left( \frac{h}{n} \right) = \left[ \frac{\lambda}{t} w \left( \frac{r_t(h)}{\lambda} \right) + \frac{\lambda}{1-t} w \left( \frac{h+r_t(h)}{\lambda} \right) \right] \mathbf{1}_{h \leq 0},$$

where  $r = r_t(h) \in [0, \lambda)$  is the unique solution of (7.10).

Let us first consider the case of  $h \leq 0$ . According to the definition (A.1) of  $\theta_{\rho_n,t}$ , and from the proof of Lemma (A.2), we know that for  $h \leq 0$ ,

$$(A.5) \quad n\theta_{\rho_n,t}(h/n) = \frac{n}{\bar{x}_n} [(1-t)u_{1-\rho_n}(1-\bar{x}_n) + tu_{1-\rho_n}(1-\bar{x}_n(1+h/n))],$$

where by (A.3),  $\bar{x}_n = \bar{x}_n(h, t)$  is the unique point in  $[1, 1/(1-\rho_n)]$  such that

$$(A.6) \quad 0 = (1-t) \log \left( 1 - (1-\rho_n) \frac{\bar{x}_n - 1}{\rho_n} \right) + t \log \left( 1 - (1-\rho_n) \left( \frac{\bar{x}_n - 1}{\rho_n} + \frac{\bar{x}_n h}{n\rho_n} \right) \right).$$

It follows that  $\lim_{n \rightarrow \infty} \bar{x}_n = 1$ . Let  $g_n(z) = (1-t) \log(1-z) + t \log(1-z+u_n)$ , with  $u_n = -(1-\rho_n)\bar{x}_n h/\lambda$ . The real  $z_n = (1-\rho_n)(\bar{x}_n - 1)/\rho_n$  is the unique solution in  $[0, 1)$  of  $g_n(z) = 0$ ,

$$0 = (1-t) \log(1-z_n) + t \log(1-z_n+u_n).$$

<sup>6</sup>For the reader's convenience, we observe that  $\frac{1}{t}\frac{\partial}{\partial h}\psi(h, x) = \log(1-\rho x(1-h)) - \log[x(1-\rho)(1-h)]$ .



Let  $\tilde{r} \in [0, 1]$  be the limit of an extracting sequence of  $(z_n)_n$ . As  $n$  goes to  $\infty$ , the above equality provides that  $r = \lambda\tilde{r}$  is the solution of

$$0 = (1 - t) \log(1 - r/\lambda) + t \log(1 - (r + h)/\lambda),$$

which is exactly (7.10). This solution is unique and therefore  $\lim_{n \rightarrow \infty} z_n = r/\lambda$ . It follows that

$$x_n = 1 + \frac{r}{n} + o\left(\frac{1}{n}\right).$$

This result provides the limit of  $n\theta_{\rho_n, t}(h/n)$  as  $n$  goes to  $\infty$ , by a Taylor expansion left to the reader.

Now, assume that  $h \geq 0$ . Following the same idea of proof, we get from (A.1) that

$$n\theta_{\rho_n, t}(h/n) = \frac{n}{\bar{x}_n} [(1 - t)u_{\rho_n}(1 - \bar{x}_n) + tu_{\rho_n}(1 - \bar{x}_n(1 + h/n))], \quad \bar{x}_n \in [1, 1/\rho_n],$$

where the real  $z_n = n(\bar{x}_n - 1)/(1 - \rho_n)$  is the unique solution of  $h_n(z) = 0$ , with

$$h_n(z) = (1 - t) \log\left(1 - \frac{\rho_n}{n} z\right) + t \log\left(1 - \frac{\rho_n}{n} \left(1 - \frac{h}{n}\right) z + \frac{h\rho_n}{n(1 - \rho_n)}\right).$$

Since  $h_n(0) \geq 0$  and for  $n$  sufficiently large  $h_n(2h) \leq 0$ , one has  $0 \leq z_n \leq 2h$ , or equivalently, for  $n$  sufficiently large

$$\bar{x}_n - 1 \leq 2h/n.$$

After few computations, this estimate implies that  $\lim_{n \rightarrow \infty} n\theta_{\rho_n, t}(h/n) = 0$ . This completes the proof of the Claim (7.9).

We now turn to the proof of Claim (7.19). We want to show that for all  $t \in (0, 1)$ ,  $h \geq 0$  and  $n \geq 2\lambda$ ,

$$n\theta_{\rho_n, t}(-h/n) \geq \frac{1}{1 - t} w\left(-\frac{1 - t}{2\lambda} h\right).$$

Since  $u_{1-\rho}(x) \geq \frac{\rho}{1-\rho} w\left(-\frac{1-\rho}{\rho} x\right)$ , the equality (A.5) implies for all  $h \geq 0$ ,

$$n\theta_{\rho_n, t}(-h/n) \geq \frac{n\rho_n}{1 - \rho_n} \frac{1}{1 - t} w\left(-\frac{1 - \rho_n}{\rho_n} (1 - \bar{x}_n(1 - h/n))\right),$$

where  $\bar{x}_n = \bar{x}_n(-h, t)$  satisfies (A.6). By the concavity of the logarithm function, (A.6) provides

$$\bar{x}_n - 1 \leq th\bar{x}_n/n,$$

and therefore

$$n\theta_{\rho_n, t}(-h/n) \geq \frac{\lambda}{1 - t} w\left(-\frac{1 - \rho_n}{\lambda} (1 - t)h\bar{x}_n\right).$$

Then the expected result (7.19) follows from the monotonicity property of  $w$  on  $\mathbb{R}_-$ , since  $\bar{x}_n \geq 1$ .  $\square$

## REFERENCES

1. R. Adamczak, *Logarithmic Sobolev inequalities and concentration of measure for convex functions and polynomial chaoses*, Bull. Pol. Acad. Sci. Math. **53** (2005), no. 2, 221–238.
2. L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 2000.
3. Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré, *Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below*, Invent. Math. **195** (2014), no. 2, 289–391. MR 3152751
4. D.P. Bertsekas and S.E. Shreve, *Stochastic optimal control*, Mathematics in Science and Engineering, vol. 139, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1978, The discrete time case.
5. S.G. Bobkov, I. Gentil, and M. Ledoux, *Hypercontractivity of Hamilton-Jacobi equations*, J. Math. Pures Appl. (9) **80** (2001), no. 7, 669–696.

6. S.G. Bobkov and F. Götze, *Exponential integrability and transportation cost related to logarithmic Sobolev inequalities*, J. Funct. Anal. **163** (1999), no. 1, 1–28.
7. A. Dembo, *Information inequalities and concentration of measure*, Ann. Probab. **25** (1997), no. 2, 927–939.
8. J.-D. Deuschel and D.W. Stroock, *Large deviations*, Pure and Applied Mathematics, vol. 137, Academic Press Inc., Boston, MA, 1989.
9. N. Gozlan, *A characterization of dimension free concentration in terms of transportation inequalities*, Ann. Probab. **37** (2009), no. 6, 2480–2498.
10. N. Gozlan and C. Léonard, *Transport inequalities. A survey*, Markov Process. Related Fields **16** (2010), no. 4, 635–736.
11. N. Gozlan, C. Roberto, and P.-M. Samson, *From concentration to logarithmic Sobolev and Poincaré inequalities*, J. Funct. Anal. **260** (2011), no. 5, 1491–1522.
12. ———, *A new characterization of Talagrand’s transport-entropy inequalities and applications*, Ann. Probab. **39** (2011), no. 3, 857–880.
13. ———, *Characterization of Talagrand’s transport-entropy inequalities in metric spaces*, Ann. Probab. **41** (2013), no. 5, 3112–3139.
14. ———, *Hamilton-Jacobi equations on metric spaces and transport-entropy inequalities*, To appear in Rev. Mat. Iberoam., 2013.
15. N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali, *Displacement convexity of entropy and related inequalities on graphs*, To appear in Probab. Theory Related Fields, 2013.
16. L. Gross, *Logarithmic Sobolev inequalities*, Amer. J. Math. **97** (1975), no. 4, 1061–1083. MR 0420249 (54 #8263)
17. J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of convex analysis*, Grundlehren Text Editions, Springer-Verlag, Berlin, 2001, Abridged version of it Convex analysis and minimization algorithms. I [Springer, Berlin, 1993; MR1261420 (95m:90001)] and it II [ibid.; MR1295240 (95m:90002)].
18. Michel Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001.
19. C. Léonard, *A saddle-point approach to the Monge-Kantorovich optimal transport problem*, ESAIM Control Optim. Calc. Var. **17** (2011), no. 3, 682–704.
20. J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2) **169** (2009), no. 3, 903–991. MR 2480619 (2010i:53068)
21. K. Marton, *A simple proof of the blowing-up lemma*, IEEE Trans. Inform. Theory **32** (1986), no. 3, 445–446.
22. ———, *Bounding  $\bar{d}$ -distance by informational divergence: a method to prove measure concentration*, Ann. Probab. **24** (1996), no. 2, 857–866.
23. ———, *A measure concentration inequality for contracting Markov chains*, Geom. Funct. Anal. **6** (1996), no. 3, 556–571.
24. B. Maurey, *Some deviation inequalities*, Geom. Funct. Anal. **1** (1991), no. 2, 188–197. MR 1097258 (92g:60024)
25. T. Mikami, *A simple proof of duality theorem for Monge-Kantorovich problem*, Kodai Math. J. **29** (2006), no. 1, 1–4.
26. F. Otto and C. Villani, *Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality*, J. Funct. Anal. **173** (2000), no. 2, 361–400.
27. K. R. Parthasarathy, *Probability measures on metric spaces*, Probability and Mathematical Statistics, No. 3, Academic Press Inc., New York, 1967.
28. S.T. Rachev and L. Rüschendorf, *Mass transportation problems. Vol. I*, Probability and its Applications (New York), Springer-Verlag, New York, 1998, Theory.
29. P.-M. Samson, *Concentration of measure inequalities for Markov chains and  $\Phi$ -mixing processes*, Ann. Probab. **28** (2000), no. 1, 416–461.
30. ———, *Concentration inequalities for convex functions on product spaces*, Stochastic inequalities and applications, Progr. Probab., vol. 56, Birkhäuser, Basel, 2003, pp. 33–52.
31. ———, *Infimum-convolution description of concentration properties of product probability measures, with applications*, Ann. Inst. H. Poincaré Probab. Statist. **43** (2007), no. 3, 321–338.
32. Y. Shu, *Hamilton jacobi equation on graphs and applications*, In preparation, 2014.
33. V. Strassen, *The existence of probability measures with given marginals*, Ann. Math. Statist. **36** (1965), 423–439.
34. M. Talagrand, *New concentration inequalities in product spaces*, Invent. Math. **126** (1996), no. 3, 505–563. MR 1419006 (99b:60030)
35. X. Tan and N. Touzi, *Optimal transportation under controlled stochastic dynamics*, Ann. Probab. **41** (2013), no. 5, 3201–3240.

36. C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.
37. ———, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009, Old and new.
38. O. Wintenberger, *Weak transport inequalities and applications to exponential and oracle inequalities*, Preprint, 2013.
39. C. Zălinescu, *Convex analysis in general vector spaces*, World Scientific Publishing Co. Inc., River Edge, NJ, 2002.

UNIVERSITÉ PARIS EST MARNE LA VALLÉE - LABORATOIRE D'ANALYSE ET DE MATHÉMATIQUES APPLIQUÉES (UMR CNRS 8050), 5 BD DESCARTES, 77454 MARNE LA VALLÉE CEDEX 2, FRANCE

UNIVERSITÉ PARIS OUEST NANTERRE LA DÉFENSE - MODAL'X, 200 AVENUE DE LA RÉPUBLIQUE 92000 NANTERRE, FRANCE

SCHOOL OF MATHEMATICS & SCHOOL OF COMPUTER SCIENCE, GEORGIA INSTITUTE OF TECHNOLOGY, ATLANTA, GA 30332

*E-mail address:*            `nathael.gozlan@univ-mlv.fr`, `cyril.roborto@math.cnrs.fr`, `paul-marie.samson@univ-mlv.fr`, `tetali@math.gatech.edu`