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. We also provide explicit examples of discrete measures satisfying the weak transport-entropy inequalities derived here.

1. INTRODUCTION

Concentration of measure phenomenon was introduced in the seventies by V. Milman [46] in his study of asymptotic geometry of Banach spaces. It was then studied in depth by many authors including Gromov [32, 31], Talagrand [62], Maurey [44], Ledoux [36, 10], Bobkov [6, 11] and many others and played a decisive role in analysis, probability and statistics in high dimensions. We refer to the monographs [37] and [16] for an overview of the field.

One classical example of such a phenomenon can be observed for the standard Gaussian measure γ_m on $\mathbb{R}^m$. It follows from the well-known Sudakov-Tsirelson-Borell isoperimetric result in Gauss space [61, 15] that if $X_1, \ldots, X_n$ are n i.i.d random vectors with law γ_m and $f : (\mathbb{R}^m)^n \to \mathbb{R}$ is a 1-Lipschitz function (with respect to the Euclidean norm), then

$$\mathbb{P}(f(X_1, \ldots, X_n) > \mathrm{med}(f) + t) \leq e^{-(t-t_o)^2/(2a)}, \qquad \forall t \geq t_o, \tag{1.1}$$

with $a = 1$ and $t_o = 0$, and where $\mathrm{med}(f)$ denotes a median of the random variable $f(X_1, \ldots, X_n)$. The remarkable feature of this inequality is that it does not depend on the sample size n. This property was used in numerous applications [37].

The standard Gaussian measure is far from being the only example of a probability distribution satisfying such a bound. In this introduction, we will say that a probability μ on some metric space (X, d) satisfies the Gaussian dimension-free concentration of measure phenomenon if (1.1) holds true, with a constant a independent of n, when the X_i's are distributed according to μ and f is a function which is 1-Lipschitz, with respect to the distance d_2 defined on X^n, by

$$d_2(x, y) = \left[\sum_{i=1}^{n} d(x_i, y_i)^2\right]^{1/2}, \qquad x, y \in X^n.$$

This is also equivalent to the following property: there exist $a > 0$ and $t_o \geq 0$ such that for all positive integers n, and all Borel sets $A \subset X^n$ such that $\mu^n(A) \geq 1/2$, it holds

$$\mu^n(A_t) \geq 1 - e^{-(t-t_o)^2/(2a)}, \qquad \forall t \geq t_o, \tag{1.2}$$

where $A_t = \{y \in X^n : \exists x \in A, d_2(x, y) \leq t\}$, is the t-enlargement of A.

Date: July 24, 2017.

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, ANR11-LBX-0023-01; 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, California).

For instance, if μ is a probability measure on $\mathbb{R}^m$, or even more generally on a smooth Riemannian manifold M equipped with its geodesic distance d and has a density of the form e^{-V}, where V is some smooth function on M such that the so-called Bakry-Émery curvature condition holds

$$\mathrm{Ric} + \mathrm{Hess}\, V \geq K\mathrm{Id}\,, \tag{1.3}$$

for some $K > 0$, then the Gaussian dimension-free concentration of measure phenomenon holds with the constant $a = K$ (a direct proof can be found in [37]).

Another very classical sufficient condition for the Gaussian concentration of measure property (1.1) is the Logarithmic Sobolev inequality introduced by Gross [33] (see also Stam [58] and Federbush [22]). If, for some $C > 0$, μ satisfies

$$\mathrm{Ent}_\mu(f^2) \leq 2C \int |\nabla f|^2 \, d\mu, \tag{1.4}$$

for all smooth functions $f : M \to \mathbb{R}$, then it satisfies (1.1) with $a = C$ (a proof of this classical result due to Herbst can be found in [37]). We recall that the entropy functional of a positive function g is defined by $\mathrm{Ent}_\mu(g) = \int g \log\left(\frac{g}{\int g\, d\mu}\right) d\mu$. Condition (1.4) - denoted $\mathbf{LSI}(C)$ in the sequel - is less restrictive since, according to the famous Bakry-Émery criterion (1.3) implies (1.4).

It turns out that Condition (1.4) can be further relaxed. Indeed, in [64], Talagrand introduced another remarkable functional inequality involving the Wasserstein distance W_2 defined, for all probability mesures μ, ν on M by

$$W_2^2(\nu, \mu) = \inf_{(X,Y)} \mathbb{E}[d^2(X,Y)], \tag{1.5}$$

where the infimum runs over all pairs of random variables (X, Y), with X distributed according to μ and Y according to ν. A probability measure μ satisfies Talagrand's transport inequality $\mathbf{T}_2(D)$ for some $D > 0$, if

$$W_2^2(\nu, \mu) \leq 2DH(\nu|\mu), \tag{1.6}$$

for all probability measures ν on M, where $H(\nu|\mu)$ denotes the relative entropy defined by $H(\nu|\mu) = \mathrm{Ent}_\mu(h)$, $h = d\nu/d\mu$, if $\nu \ll \mu$ (i.e., ν absolutely continuous with respect to μ) and $+\infty$ otherwise. A nice argument first discovered by Marton [41] shows that (1.6) is a sufficient condition for the Gaussian dimension-free concentration property (1.1) with $a = D$. One crucial ingredient to derive dimension-free concentration from $\mathbf{T}_2$ is the tensorization property enjoyed by this inequality (the same property holds for $\mathbf{LSI}$): if μ satisfies $\mathbf{T}_2(C)$, then for any positive integer n, the product measure μ^n also satisfies $\mathbf{T}_2(C)$. Condition (1.6) is again an improvement upon Condition (1.4), since it was proved by Otto and Villani [49] (see [7] for an alternative proof and [28] and the references therein for extensions to more general spaces) that (1.4) implies (1.6) with $D = C$. It was then shown by the first author [23] that Condition (1.6) was not only sufficient but also necessary for Gaussian dimension-free concentration. More precisely, if (1.1) holds true with some a (and all n), then μ satisfies $\mathbf{T}_2(a)$.

One of the main motivations behind this work, and a few satellite papers by the same authors and Y. Shu [30, 57, 29], is to understand what can replace each term in the chain of implications:

$$(1.3) \Rightarrow (1.4) \Rightarrow (1.6) \Leftrightarrow (1.1)$$

in a *discrete* setting (for instance, when the space is a graph, finite or otherwise).

While several useful variants of the logarithmic Sobolev inequality are identified in discrete settings (involving different natural discrete gradients, see *e.g.* [52, 12]), the other terms are far from well understood.

After the works by Lott-Villani [39] and Sturm [60] extending (1.3) to non-smooth geodesic spaces through convexity properties of the entropy functional on the space of probability measures equipped with the Wasserstein distance W_2, the question of generalizing the Bakry-Émery condition in a discrete setting attracted in recent years a lot of attention. We refer to the works by Ollivier [47], Bonciocat-Sturm [14], Ollivier-Villani [48], Erbar-Maas [21], Hillion [34] and the work [30] by the authors for different attempts to give a meaning to the notion of "discrete curvature".

In the present paper, the focus is put on the rightmost terms of our chain of implications: namely, our purpose is to find out what type of dimension-free concentration results we can hope for in a discrete setting and what type of transport inequalities can be related to it. At this stage, it is worth noting that, unfortunately, Talagrand's inequality is *never* satisfied in discrete (except of course by a Dirac mass). For instance, it is proven in full generality in [28], that if μ is a probability measure on a metric space (X, d) which satisfies $\mathbf{T}_2$, then its support is connected. It follows from the equivalence (1.6) $\Leftrightarrow$ (1.1), that Gaussian dimension-free concentration is also never true in discrete settings.

One thus looks for a transport-cost sufficiently weaker than W_2^2, to allow discrete measures to satisfy the related transport inequality, but sufficiently strong to make the transport inequality stable under tensor products. A natural candidate would be the W_1 distance: $W_1(\nu,\mu) := \inf\{\mathbb{E}[d(X,Y)] : \mathrm{Law}(X) = \mu, \mathrm{Law}(Y) = \nu\}$. Although transport inequalities involving W_1^2 instead of W_2^2 (the so called $\mathbf{T}_1$ inequalities) make perfect sense in discrete settings (see Bobkov-Götze [8], Djelout-Guillin-Wu [19], Bobkov-Houdré-Tetali [9]), these inequalities tensorize only with a constant depending on the dimension! So the W_1 distance does not fulfill the second requirement.

The present paper is devoted to the study of a family of weak transport costs, one typical element of which is the following weak version of the cost W_2^2 defined as follows. If μ and ν are probability measures on a metric space (X, d), one defines the weak cost $\widetilde{\mathcal{T}}_2(\nu|\mu)$ as follows:

$$\widetilde{\mathcal{T}}_2(\nu|\mu) = \inf_{(X,Y)} \mathbb{E}\left[\,\mathbb{E}[d(X,Y)|X]^2\,\right],$$

where again the infimum runs over all pairs (X, Y) of random variables such that X follows the law μ and Y the law ν. Jensen's inequality immediately shows that

$$W_1^2(\nu,\mu) \le \widetilde{\mathcal{T}}_2(\nu|\mu) \le W_2^2(\nu,\mu).$$

Two weak versions of Talagrand's inequality are naturally associated to this cost: a probability μ is said to satisfy $\widetilde{\mathbf{T}}_2^-(C)$ for some $C > 0$ if

$$\widetilde{\mathcal{T}}_2(\mu|\nu) \le CH(\nu|\mu), \qquad \forall \nu$$

and to satisfy $\widetilde{\mathbf{T}}_2^+(C)$ if

$$\widetilde{\mathcal{T}}_2(\nu|\mu) \le CH(\nu|\mu), \qquad \forall \nu.$$

Since $\widetilde{\mathcal{T}}_2$ is not symmetric these two inequalities are not equivalent in general. Both are of course implied by the usual $\mathbf{T}_2(C)$ inequality. As we shall see in Theorem 5.1, which is one of our main results, a probability measure μ on X satisfies the two inequalities $\widetilde{\mathbf{T}}_2^\pm(C)$, for some $C > 0$, if and only if it satisfies the following dimension-free concentration of measure property: for all positive integers n and all sets $A \subset X^n$ such that $\mu^n(A) > 0$, it holds

$$\mu^n(\widetilde{A}_t) \ge 1 - \frac{1}{\mu^n(A)} e^{-t^2/D}, \qquad \forall t \ge 0, \tag{1.7}$$

for some D related to C. In this concentration inequality, the enlargement $\widetilde{A}_t$ of A is defined as follows

$$\widetilde{A}_t = \{y \in X^n : \exists p \in \mathcal{P}(X^n) \text{ with } p(A) = 1 \text{ such that } \sum_{i=1}^n \left(\int_A d(x_i, y_i)\, p(dx)\right)^2 \le t^2\},$$

where $\mathcal{P}(X^n)$ is the set of all Borel probability measures on X^n. Taking $p = \delta_x$ with $x \in A$, we see immediately that $A_t \subset \widetilde{A}_t$ and therefore (1.7) is less demanding than (1.2).

Before going further into the presentation of our results, let us make some bibliographical comments on these weak transport costs and on the concentration property (1.7). First of all, this way of enlarging sets first appeared in the papers [62, 63] by Talagrand, in the particular case where $d(x, y) = \mathbf{1}_{x\neq y}$ is the Hamming distance (see [62, Theorem 4.1.1] and [63, (1.2)]). It was shown by Talagrand that *any* probability measure μ on a Polish space X satisfies the concentration inequality (1.7) with some universal constant D (and with the Hamming distance). This deep result known as Talagrand's convex hull concentration inequality has had a lot of interesting applications in probability theory and combinatorics [62, 37, 2]. The result was given another proof by Marton in

[43], where she introduced (again with d being the Hamming distance) the weak transport cost $\widetilde{\mathcal{T}}_2$ (denoted $\bar{d}_2$ in her work) and proved that any probability measure μ satisfies

$$\widetilde{\mathcal{T}}_2(\nu_1|\nu_2)^{1/2} \leq (2H(\nu_1|\mu))^{1/2} + (2H(\nu_2|\mu))^{1/2},$$

for all probability measures ν_1, ν_2. Then she proved the tensorization property for this transport inequality and derived from it, using an argument that will be recalled in Section 5, Talagrand's concentration result. A similar strategy was then developed by Dembo in [17] in order to recover the sharp form of other concentration results by Talagrand. Finally the third-named author extended (in [53]) the tensorization technique of Marton to some classes of dependent random variables; in this work he improved Marton's transport inequality to recover yet another sharp concentration inequality by Talagrand (discovered in [63]) related to deviation inequalities for empirical processes. Besides the Hamming case, almost nothing is known on the inequalities $\widetilde{\mathbf{T}}_2^{\pm}$. Note that Marton's result shows at least that any probability measure on a bounded metric space (X, d) (for instance a finite graph equipped with graph distance) satisfies $\widetilde{\mathbf{T}}_2^{\pm}(C)$ for $C = 2\mathrm{Diam}(X)$, but the optimal constant C can be much smaller. Note also that if our primary motivation was to consider these inequalities on discrete spaces, the continuous case is also of interest since the inequalities $\widetilde{\mathbf{T}}_2^{\pm}$ could be a good substitute for probability measures not satisfying the usual $\mathbf{T}_2$. The aim of the paper is thus to provide different tools that can be useful in the study of these weak transport cost inequalities and to exhibit some new examples of such inequalities (mostly on unbounded spaces).

The main new tool we introduce is a version of the Kantorovich duality theorem suitable for the weak transport cost $\widetilde{\mathcal{T}}_2$. Actually this duality result holds for a large family of transport costs that we shall now describe. To each cost function $c : X \times \mathcal{P}(X) \to [0, \infty]$ (where $\mathcal{P}(X)$ is the set of Borel probability measures on X), we associate the optimal transport cost $\mathcal{T}_c$ defined, for all probability measures μ, ν on X, by

$$\mathcal{T}_c(\nu|\mu) = \inf_p \int c(x, p_x)\, \mu(dx), \tag{1.8}$$

where the infimum runs over the set of all probability kernels $p : X \to \mathcal{P}(X) : x \mapsto p_x(\,\cdot\,)$ such that $\mu p = \nu$. Note that the usual cost W_2^2 corresponds to $c(x, p) = \int d^2(x, y)\, p(dy)$ and the weak cost $\widetilde{\mathcal{T}}_2$ to $c(x, p) = \left(\int d(x, y)\, p(dy)\right)^2$. Under some easily satisfiable technical assumptions on c (an important one being that c be *convex* with respect to its second variable p, *i.e.* $c(x, (1-\lambda)p_1+\lambda p_2) \leq (1-\lambda)c(x, p_1) + \lambda c(x, p_2)$ for all $\lambda \in [0, 1]$, $p_1, p_2 \in \mathcal{P}(X)$, $x \in X$), we prove in Theorem 9.6 that

$$\mathcal{T}_c(\nu|\mu) = \sup \left\{ \int R_c\varphi(x)\, \mu(dx) - \int \varphi(y)\, \nu(dy) \right\},$$

where the supremum runs over the set of bounded continuous functions, and

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

Note that, when $c(x, p) = \int d^2(x, y) p(dy)$, then $R_c\varphi(x) = \inf_{y \in X}\{\varphi(y) + d^2(x, y)\}$ and the result reduces to the classical Kantorovich duality for W_2^2 (see *e.g.* [66, 67]). Up to our best knowledge, this class of cost functionals has not been considered before in the literature on Optimal Transport, but we think that it may find interesting applications in this field. For example, denoting by $\overline{\mathcal{T}}_r$ the weak cost associated to the cost function $c(x, p) = \|x - \int y\, p(dy)\|^r$ defined on $\mathbb{R}^m \times \mathcal{P}(\mathbb{R}^m)$ where $\|\cdot\|$ is some norm on $\mathbb{R}^m$, it turns out that the duality formula for $\overline{\mathcal{T}}_1$ immediately gives back a well-known result by Strassen [59] about the existence of martingales with given marginals. This is detailed in Section 3.

The paper is organized as follows.

Section 2 introduces a general definition of optimal transport costs and presents in detail three particular families of costs (all variants of Marton's costs $\widetilde{\mathcal{T}}_2$ defined above) which will play a central role in the rest of the paper. In particular, we state a Kantorovich duality formula for each of these transport costs.

Section 3 is dedicated to the proof of Strassen's theorem on the existence of martingales with given marginals.

Section 4 introduces the general definition of transport-entropy inequalities (involving general transport costs of the form (1.8)) and presents their basic properties such as their dual formulation and their tensorization.

Section 5 deals with the links between concentration of measure and transport-entropy inequalities. We recall in particular the argument due to Marton that enables to deduce concentration estimates from transport-entropy inequalities. We also extend to this general framework a result by the first author and show that in great generality dimension-free concentration gives back transport-entropy inequalities. In particular, we give a characterization (in terms of a transport-entropy inequality involving the cost $\overline{\mathcal{T}}_2$ defined above) of dimension-free Gaussian concentration (1.1) restricted to Lipschitz convex (or concave) functions.

Section 6 recalls the universal transport-entropy inequalities developed by Marton [43, 42], Dembo [17] and Samson [54, 55] in order to recover some of Talagrand's concentration inequalities for product measures. These transport-entropy inequalities are seminal examples that motivated this work. We take advantage of our duality theorems to revisit and bring some simplifications in the proof of [55].

Section 7 contains examples of sharp transport-entropy inequalities for Bernoulli and binomial laws obtained by tensorization and projection arguments. Following this, by weak convergence of the binomial towards the Poisson distribution, we introduce a new weak transport-entropy inequality for the Poisson distribution with an optimal cost function.

Concerning examples of transport-entropy inequality for discrete measures, let us mention that new weak transport-entropy inequalities for the uniform measure or the Ewens distribution on the symmetric group have been recently settled by the third named author in [56], with extensions to "locally acting" groups of permutations. These transport inequalities give back the concentration results on the symmetric group obtained by Talagrand in [62] and Luczak-McDiarmid in [40].

In Section 8 we show the equivalence between transport-entropy inequalities involving the transport cost $\overline{\mathcal{T}}_2$ and the logarithmic-Sobolev inequality restricted to the class of log-convex or log-concave functions. This enables us to get other examples for these transport inequalities.

Finally, Section 9 contains the proof of our general Kantorovich duality result, Theorem 9.6, for a transport cost of the form (1.8).

Contents

2. Optimal transport costs and duality

In this section, we introduce a general class of optimal transport costs and describe an associated Kantorovich type duality formula.

2.1. **Notations.**

Throughout the paper (X, d) is a complete separable metric space. The Borel σ-field will be denoted by $\mathcal{B}$. The space of all Borel probability measures on X is denoted by $\mathcal{P}(X)$.

Let $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function such that

$$\gamma(u+v) \leq C(\gamma(u)+\gamma(v)), \qquad u, v \in \mathbb{R}_+, \tag{2.1}$$

for some constant C, then we set

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

for some $x_o \in X$. Note that, thanks to Condition (2.1), the definition of $\mathcal{P}_\gamma(X)$ is independent of x_o. In the specific cases where $\gamma_r(u) := u^r$, $u \geq 0$, $r \geq 0$, we use the simpler notation $\mathcal{P}_r(X) := \mathcal{P}_{\gamma_r}(X)$. Observe that since $\gamma_0(u) = 1, u \geq 0$, one has $\mathcal{P}_0(X) = \mathcal{P}_{\gamma_0}(X) = \mathcal{P}(X)$.

We also denote by $\Phi_\gamma(X)$ (resp. $\Phi_{\gamma,b}(X)$) the set of continuous (resp. continuous and bounded from below) functions $\varphi : X \to \mathbb{R}$ satisfying the growth condition

$$|\varphi(x)| \leq a + b\gamma(d(x, x_o)), \qquad \forall x \in X, \tag{2.2}$$

for some $a, b \geq 0$ and some (hence all, thanks to (2.1)) $x_o \in X$. With our notations $\Phi_{\gamma_0}(X) = \Phi_{\gamma_0,b}(X)$ is the set of continuous bounded real functions on X usually denoted by $\mathcal{C}_b(X)$.

The space $\Phi_\gamma(X \times X)$ and $\mathcal{P}_\gamma(X \times X)$ are defined accordingly. The set $\Phi_\gamma(X \times X)$ is the set of all continuous functions $\varphi : X \times X \to \mathbb{R}$ such that there exist $a, b \geq 0$ and $x_o \in X$ such that

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

and similarly

$$\mathcal{P}_\gamma(X \times X) = \left\{\pi \in \mathcal{P}(X \times X); \int \gamma(d(x,x_o)) + \gamma(d(y,x_o))\,\pi(dxdy) < \infty\right\}.$$

The space $\mathcal{P}_\gamma(X)$ will always be equipped with the σ-field $\mathcal{F}_\gamma$ generated by the maps

$$\mathcal{P}_\gamma(X) \to [0,1] : \nu \mapsto \nu(A),$$

where A is a Borel set of X. In particular, one says that $p : X \to \mathcal{P}_\gamma(X) : x \mapsto p_x$ is a kernel if it is measurable with respect to the Borel σ-field $\mathcal{B}$ on X and the σ-field $\mathcal{F}_\gamma$ on $\mathcal{P}_\gamma(X)$. This amounts to requiring that, for all $A \in \mathcal{B}$, the map $X \to [0,1] : x \mapsto p_x(A)$ be Borel measurable.

2.2. **Costs functions, couplings and weak optimal transport costs.** Let $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ be a fixed continuous function satisfying (2.1). In this paper, a *cost function* will be a measurable function $c\colon X \times \mathcal{P}_\gamma(X) \to [0,\infty]$ and we set, for all $\pi \in \mathcal{P}_\gamma(X \times X)$,

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

where π_1 is the first marginal of π and $x \mapsto p_x$ the (π_1-almost everywhere, uniquely determined) probability kernel 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 π_1 almost all $x \in X$ and thus the preceding definition makes sense.

Given two probability measures μ and ν 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* π whose first marginal is μ and whose second marginal is ν. Note also that if, $\mu,\nu \in \mathcal{P}_\gamma(X)$, then $\Pi(\mu,\nu) \subset \mathcal{P}_\gamma(X \times X)$.

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

Definition 2.3. *Let $c\colon X \times \mathcal{P}_\gamma(X) \to [0,\infty]$ and $\mu,\nu \in \mathcal{P}_\gamma(X)$. The optimal transport cost $\mathcal{T}_c(\nu|\mu)$ between μ and ν is defined by*

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

Let us first remark that optimal transport costs in the classical sense (see e.g. [66, 67]) enter the framework of this definition. Namely, if $\omega : X \times X \to [0,\infty]$ is some measurable cost function, and $c(x,p) = \int \omega(x,y)\,p(dy)$, for all $x \in X$ and $p \in \mathcal{P}(X)$, then it is clear that

$$\mathcal{T}_c(\nu|\mu) = \inf\left\{\iint \omega(x,y)\,\pi(dxdy) : \pi \in \Pi(\mu,\nu)\right\},$$

which is the usual optimal transport cost related to the cost function ω. In the sequel, we will denote by $\mathcal{T}_\omega(\nu,\mu)$ the usual Monge-Kantorovich optimal transport cost, defined by the right hand side above. One sees that while in the usual definition every elementary transport of mass from μ to ν represented by p_x is penalized by its mean cost $\int \omega(x,y)\,p_x(dy)$, our definition allows other types of penalization. See Section 2.4 below for some examples.

2.3. **A Kantorovich type duality.** If $\omega : X \times X \to [0,\infty]$ is lower semi-continuous, then according to the well-known Kantorovich duality theorem (see for instance [67, Theorem 5.10]), it holds

$$\mathcal{T}_\omega(\nu,\mu) = \sup \left\{ \int \psi(x)\,\mu(dx) - \int \varphi(y)\,\nu(dy) \right\},$$

where the supremum runs over the class of pairs (ψ,φ) of bounded continuous functions on X such that

$$\psi(x) - \varphi(y) \leq \omega(x,y), \qquad \forall x,y \in X.$$

A classical and simple argument shows that one can always replace ψ by the function $Q_\omega\varphi$ defined by

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

Therefore, the duality formula above can be restated as follows

$$\mathcal{T}_\omega(\nu,\mu) = \sup \left\{ \int Q_\omega\varphi(x)\,\mu(dx) - \int \varphi(y)\,\nu(dy) \right\},$$

where the supremum runs over the class of bounded continuous functions φ. In case the function $Q_\omega\varphi$ is not measurable, then we understand $\int Q_\omega\varphi(x)\mu(dx)$ as the integral with respect to the inner measure μ_* induced by μ. Recall that if $g : X \to \mathbb{R}$ is a function bounded from below, then

$$\int g(x)\,\mu_*(dx) = \sup \int f(x)\,\mu(dx),$$

where the supremum runs over the set of bounded *measurable* functions f such that $f \leq g$.

Under some semi-continuity and convexity assumptions on the cost function c, this duality formula generalizes to our optimal transport costs in the sense of Definition 2.3. The duality property is described in the following definition.

Definition 2.4. *Let $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function satisfying (2.1) and $c : X \times \mathcal{P}_\gamma(X) \to [0,+\infty]$ be a measurable cost function. One says that* duality holds for the cost function c, *if for all $\mu,\nu \in \mathcal{P}_\gamma(X)$, it 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

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

Observe that the equality in the above definition has to be understood in $[0,+\infty]$. Also we stress that, by the very definitions of $\mathcal{T}_c$ and $R_c\varphi$, the following inequality always holds $\sup_{\varphi\in\Phi_{\gamma,b}(X)} \left\{\int R_c\varphi(x)\,\mu_*(dx) - \int \varphi(y)\,\nu(dy)\right\} \leq \mathcal{T}_c(\nu|\mu)$ so that only one direction is difficult to prove.

Section 9 is devoted to the proof of a general result showing that duality holds under mild regularity conditions on c. Among these conditions, the main requirement is that c is convex with respect to the variable p. We refer to Theorem 9.6 for a precise statement. Since we do not know whether the conditions of Theorem 9.6 are minimal, we prefer to postpone its statement to Section 9 and to focus on particular families of cost functions (which are especially relevant for the applications we have in mind) for which the duality holds.

2.4. **Particular cases.** As we already observed, if $\omega : X\times X \to [0,\infty]$ is a measurable function and $c(x,p) = \int \omega(x,y)\,p(dy)$, $x \in X, p \in \mathcal{P}(X)$, then the associated optimal transport cost corresponds to the usual Monge-Kantorovich optimal transport cost $\mathcal{T}_\omega$ defined by

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

Among these costs a popular choice consists of taking, $\omega(x,y) = \alpha(d(x,y))$, for $x,y \in X$, where $\alpha : \mathbb{R}^+ \to \mathbb{R}^+$ is a convex function.

The simple idea that leads from this classical family of cost functions to the family of cost functions described below, is to weaken c by applying Jensen's inequality:

$$c(x,p) = \int \alpha\left(d(x,y)\right)\, p(dy) \geq \alpha\left(\int d(x,y)\, p(dy)\right) := \tilde{c}(x,p).$$

Cost functions of the form $\tilde{c}$ as above appeared, in the particular case of the Hamming distance, in papers by Marton [43, 42], Dembo [17], Samson [53, 54, 55] in their studies of transport-type inequalities related to Talagrand's universal concentration inequalities for independent random variables. See Section 6 for more information on the topic.

2.4.1. *Marton's cost functions.* Fix a lower-semicontinuous function $\delta : \mathbb{R}_+ \to \mathbb{R}_+$ and a convex function $\alpha : \mathbb{R}_+ \to [0,+\infty]$. The optimal transport cost associated to the cost function

$$c(x,p) = \alpha\left(\int \delta(d(x,y))\, p(dy)\right), \qquad x \in X, \qquad p \in \mathcal{P}_\gamma(X), \tag{2.5}$$

will be denoted by $\widetilde{\mathcal{T}}_\alpha$ and is defined by

$$\widetilde{\mathcal{T}}_\alpha(\nu|\mu) = \inf_{\pi\in\Pi(\mu,\nu)} \int \alpha\left(\int \delta(d(x,y))\, p_x(dy)\right) \mu(dx)\,, \tag{2.6}$$

where $x \mapsto p_x$ is the probability kernel defined as usual by $\pi(dxdy) = \mu(dx)p_x(dy)$. We will refer to this family of cost functions / optimal transport costs as *Marton's costs* since they were first considered in [43] for $\delta(u) := \mathbf{1}_{u\neq 0}$, $u \geq 0$, and α being the quadratic function, in this case $\delta(d(x,y)) = \mathbf{1}_{x\neq y}, x, y \in X$, and therefore $c(x,p) = \left(\int \mathbf{1}_{x\neq y}\, p(dy)\right)^2 = p(X \setminus \{x\})^2$.

Note that, in general, $\widetilde{\mathcal{T}}_\alpha$ is not symmetric in μ, ν. Moreover, as we already observed above, if $\omega(x,y) = \alpha(\delta(d(x,y)))$, then by Jensen's inequality,

$$\widetilde{\mathcal{T}}_\alpha(\nu|\mu) \leq \mathcal{T}_\omega(\nu,\mu).$$

Finally, using probabilistic notations, one has

$$\widetilde{\mathcal{T}}_\alpha(\nu|\mu) = \inf_{(X,Y)} \mathbb{E}\Big[\alpha\big(\mathbb{E}\left[\delta(d(X,Y))\,|\,X\right]\big)\Big],$$

where the infimum runs over the set of all pairs of random variables (X,Y), where X has law μ and Y has law ν. The following result gives sufficient conditions for duality for Marton's costs.

Theorem 2.7. *Assume either that*

- *(X,d) is a complete separable metric space, $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ is a convex continuous function, $\delta : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous and $\gamma = \delta$ satisfies (2.1),*
- *or (X,d) is either a compact space or a countable set of isolated points, $\alpha : \mathbb{R}_+ \to [0,+\infty]$ is a convex lower-semicontinuous function, $\delta : \mathbb{R}_+ \to \mathbb{R}_+$ is lower-semicontinuous and $\gamma = \gamma_0$.*

Then, duality holds for the cost function c defined in (2.5). More precisely,

$$\widetilde{\mathcal{T}}_\alpha(\nu|\mu) = \sup_{\varphi\in\Phi_{\gamma,b}(X)} \left\{\int \widetilde{Q}_\alpha\varphi(x)\, \mu(dx) - \int \varphi(y)\, \nu(dy)\right\},\ \mu,\nu \in \mathcal{P}_\gamma(X), \tag{2.8}$$

where

$$\widetilde{Q}_\alpha\varphi(x) = \inf_{p\in\mathcal{P}_\gamma(X)} \left\{\int \varphi(y)\, p(dy) + \alpha\left(\int \delta(d(x,y))\, p(dy)\right)\right\},$$

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

We observe that, anticipating the present paper, the duality formula (2.8) was already put to use in [30], in connection with displacement convexity of the relative entropy functional on graphs.

2.4.2. *A barycentric variant of Marton's cost functions.* When $X \subset \mathbb{R}^m$ (equipped with an arbitrary norm $\|\cdot\|$) is a closed set, a variant of Marton's costs functions is obtained by choosing

$$c(x,p) = \theta\left(x - \int y\, p(dy)\right), \qquad x \in X, \qquad p \in \mathcal{P}_1(X), \tag{2.9}$$

where $\theta : \mathbb{R}^m \to [0,\infty]$ is a lower-semicontinuous convex function (for such a cost we choose $\gamma(u) = \gamma_1(u) = u$, $u \geq 0$). The corresponding transport cost is denoted by $\overline{\mathcal{T}}_\theta$ and defined by

$$\overline{\mathcal{T}}_\theta(\nu|\mu) = \inf_{\pi \in \Pi(\mu,\nu)} \int \theta\left(x - \int y\, p_x(dy)\right) \mu(dx). \tag{2.10}$$

We use the notation $\overline{\mathcal{T}}_\theta$ with a *bar* in reference to the *barycenter* entering its definition.

Using probabilistic notations, we have the following alternative definition

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

where the infimum runs over the set of all pairs of random variables (X,Y), with X having law μ and Y having law ν. Moreover, if $\omega(x,y) = \alpha(\|x-y\|)$, $x,y \in \mathbb{R}^m$, where $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ is convex, and $\theta(u) = \alpha(\|u\|)$, $u \in \mathbb{R}^m$, then the following holds:

$$\mathcal{T}_\omega(\nu,\mu) \geq \widetilde{\mathcal{T}}_\alpha(\nu|\mu) \geq \overline{\mathcal{T}}_\theta(\nu|\mu).$$

As we shall see below, this 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 3).

Duality for this family of costs functions is established in the following result. Note that for the "bar" transport cost, the duality formula for $\overline{\mathcal{T}}_\theta$ can be expressed using only convex functions. This fact will repeatedly be used in the applications.

Theorem 2.11. *Let $X \subset \mathbb{R}^m$ be a closed subset of $\mathbb{R}^m$ equipped with a norm $\|\cdot\|$ and $\theta\colon \mathbb{R}^m \to \mathbb{R}_+$ 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 duality holds for the cost function defined in (2.9). More precisely:*

(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 $\overline{\mathcal{T}}_\theta(\nu|\mu)$, 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 $h : \mathbb{R}^m \to \mathbb{R}$ 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

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

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

2.4.3. *Samson's cost functions.* Let $\beta : \mathbb{R}_+ \to [0,+\infty]$ be a lower-semicontinuous convex function and let $\delta : \mathbb{R}_+ \to \mathbb{R}_+$ be a lower-semicontinuous function with $\delta(0) = 0$. Let μ_0 be a reference probability measure on X. The choice

$$c(x,p) = \int \beta\left(\delta(d(x,y))\frac{dp}{d\mu_0}(y)\right)\mu_0(dy), \qquad x \in X, \tag{2.12}$$

if $p \in \mathcal{P}(X)$ is absolutely continuous with respect to μ_0 on $X \setminus \{x\}$, and $c(x,p) = +\infty$ otherwise, yields the family of weak transport $\widehat{\mathcal{T}}_\beta$ defined by

$$\widehat{\mathcal{T}}_\beta(\nu|\mu) = \inf_{\pi\in\Pi(\mu,\nu)} \iint \beta\left(\delta(d(x,y))\frac{dp_x}{d\mu_0}(dy)\right)\mu_0(dy)\,\mu(dx), \tag{2.13}$$

for all measures $\mu,\nu \in \mathcal{P}(X)$ (in this last expression we assume that the set of $\pi(dxdy) = \mu(dx)p_x(dy) \in \Pi(\mu,\nu)$ such that, μ almost surely, p_x is absolutely continuous with respect to μ_0 on $X \setminus \{x\}$, is not empty). Cost functions of this type were introduced by the third-named author in [55] with $\delta(u) = \mathbf{1}_{u\neq 0}, u \geq 0$, and therefore $\delta(d(x,y)) = \mathbf{1}_{x\neq y}$, $x, y \in X$.

Again, if $\beta = \alpha$ is convex, then Jensen inequality gives

$$\widetilde{\mathcal{T}}_\beta(\nu|\mu) \leq \widehat{\mathcal{T}}_\beta(\nu|\mu),$$

but there is no clear comparison between $\widehat{\mathcal{T}}_\beta(\nu|\mu)$ and $\mathcal{T}_\omega(\nu|\mu)$ with $\omega(x,y) = \beta(\delta(d(x,y)))$, $x, y \in X$.

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

Theorem 2.14. *Let (X,d) be a compact metric space or a countable set of isolated points. Let $\beta : \mathbb{R}_+ \to [0,+\infty]$ be a lower-semicontinuous convex function with $\beta(0) = 0$ and $\lim_{x\to\infty}\beta(x)/x = +\infty$. Assume that $\delta : \mathbb{R}_+ \to \mathbb{R}_+$ is lower-semicontinuous with $\delta(0) = 0$ and $\delta(u) > 0$ for all $u > 0$. Let μ_0 be a reference probability measure on X. Then duality holds for the cost function defined in (2.12) (with $\gamma = \gamma_0$). More precisely, for all $\mu,\nu \in \mathcal{P}(X)$ absolutely continuous with respect to μ_0, it holds*

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

where for $x \in X$ and $\varphi \in \mathcal{C}_b(X)$,

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

2.4.4. *Notation.* We end this section by introducing notations for the optimal transport costs related to power functions. When $\alpha(x) = x^p$, $x \geq 0$, $p > 0$, we will use the notation $\mathcal{T}_p$ and $\widetilde{\mathcal{T}}_p$ 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 $\overline{\mathcal{T}}_p$.

3. Proof of a result by Strassen

In this short section, we show that the transport cost $\overline{\mathcal{T}}_\theta$ can be used to recover an old result by Strassen [59] 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 μ is dominated by ν 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[1] $f : \mathbb{R}^m \to \mathbb{R}$. Note that, in particular, this implies that $\int f\,d\mu = \int f\,d\nu$ for all affine maps $f : \mathbb{R}^m \to \mathbb{R}$.

[1] Note that since $\mu,\nu \in \mathcal{P}_1(\mathbb{R}^m)$, any affine map is integrable with respect to μ and ν. 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 μ and ν makes sense.

It is not difficult to check that $\mu \preceq_C \nu$ if and only if $\int f\,d\mu \le \int f\,d\nu$ for all 1-Lipschitz and convex $f:\mathbb{R}^m\to\mathbb{R}$ bounded from below[2].

The following result goes back at least to the work of Strassen [59, Theorem 2].

Theorem 3.1. *Let $\mu,\nu\in\mathcal{P}_1(\mathbb{R}^m)$; there exists a martingale (X,Y), where X follows the law μ and Y the law ν if and only if $\mu \preceq_C \nu$.*

Below we obtain Strassen's theorem as a consequence of the duality formula for the cost $\overline{\mathcal{T}}_1$ given in the following proposition.

Proposition 3.2. *For all $\mu,\nu\in\mathcal{P}_1(\mathbb{R}^m)$,*

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

Proof. We already know from Item (3) of Theorem 2.11 that for all $\mu,\nu\in\mathcal{P}_1(\mathbb{R}^m)$ it holds

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

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 $\varphi:\mathbb{R}^m\to\mathbb{R}$ is convex and bounded from below, so is $Q_1\varphi\colon\mathbb{R}^m\to\mathbb{R}$. Being an infimum of 1-Lipschitz functions, $Q_1\varphi$ is itself 1-Lipschitz. Moreover, if $\psi:\mathbb{R}^m\to\mathbb{R}$ is some 1-Lipschitz convex function, then $Q_1\psi=\psi$; namely, for all $x\in\mathbb{R}^m$, one has

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

From these considerations, we conclude that

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

where the first inequality follows by letting $\psi=Q_1\varphi$ which satisfies $\psi\le\varphi$. This concludes the proof. □

Proof of Theorem 3.1. If $\pi\in\Pi(\mu,\nu)$ denotes the law of (X,Y), the condition that (X,Y) is a martingale is expressed by

$$\int y\,p_x(dy) = x,\qquad \text{for } \mu \text{ almost every } x\in\mathbb{R}^m. \tag{3.3}$$

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 (3.3) if and only if $\overline{\mathcal{T}}_1(\nu|\mu)=0$. Note that for the converse, we use the fact that there exists an optimal π in the definition of $\overline{\mathcal{T}}_1(\nu|\mu)$. This easily follows from Corollary 9.12 (which applies since the related cost function satisfies assumption (C_1) as shows the proof of Theorem 2.11). Since, by Proposition 3.2,

$$\overline{\mathcal{T}}_1(\nu|\mu) = \sup\left\{\int f\,d\mu - \int f\,d\nu; f\colon\mathbb{R}^m\to\mathbb{R}, 1-\text{Lipschitz, convex and bounded below}\right\},$$

the expected result follows. □

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

Remark 3.4. *Let us note that we obtained in fact the following slightly more general result: Let $\varepsilon > 0$; two probability measures $\mu, \nu \in \mathcal{P}_1(\mathbb{R}^m)$ satisfy $\int f\, d\mu \leq \int f\, d\nu + \varepsilon$, for all 1-Lipschitz convex functions $f : \mathbb{R}^m \to \mathbb{R}$, if and only if there exists a pair (X, Y) of random variables, with X of law μ and Y of law ν, such that*

$$\mathbb{E}[\|X - \mathbb{E}[Y|X]\|] \leq \varepsilon.$$

4. Transport-entropy inequalities: definitions, tensorization, and dual formulation

In this section, we introduce a general notion of transport-entropy inequalities of Talagrand-type and investigate them.

4.1. Definitions.

We recall that if μ, ν are two probability measures on some space X, the relative entropy of ν with respect to μ 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, ones sets $H(\nu|\mu) = +\infty$.

Definition 4.1 (Transport-entropy inequalities $\mathbf{T}_c(a_1, a_2)$ and $\mathbf{T}_c(b)$)**.**
Let $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function satisfying (2.1). Let $c\colon X \times \mathcal{P}_\gamma(X) \to [0, \infty]$ be a measurable cost function and $\mu \in \mathcal{P}_\gamma(X)$.

- *The probability measure μ is said to satisfy $\mathbf{T}_c(a_1, a_2)$, for some $a_1, a_2 > 0$ if*

$$\mathcal{T}_c(\nu_1|\nu_2) \leq a_1 H(\nu_1|\mu) + a_2 H(\nu_2|\mu), \qquad \forall \nu_1, \nu_2 \in \mathcal{P}_\gamma(X). \tag{4.2}$$

- *The probability measure μ is said to satisfy $\mathbf{T}_c^+(b)$ for some $b > 0$, if*

$$\mathcal{T}_c(\nu|\mu) \leq bH(\nu|\mu), \qquad \forall \nu \in \mathcal{P}_\gamma(X). \tag{4.3}$$

- *The probability measure μ is said to satisfy $\mathbf{T}_c^-(b)$ for some $b > 0$, if*

$$\mathcal{T}_c(\mu|\nu) \leq bH(\nu|\mu), \qquad \forall \nu \in \mathcal{P}_\gamma(X). \tag{4.4}$$

For the specific transport costs $\widetilde{\mathcal{T}}_p$ and $\overline{\mathcal{T}}_p$ introduced in Section 2.4.4 we may use the corresponding notations $\widetilde{\mathbf{T}}_p(a_1, a_2)$, $\widetilde{\mathbf{T}}_p^\pm(b)$, 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 $c(x, p) = \int \omega(x, y)\, p(dy)$, (4.3) and (4.4) give back the usual transport-entropy inequalities of Talagrand type (see [37], [67] or [24] for a general introduction on the subject). Also, we observe that $\mathbf{T}_c(a_1, 0)$ or $\mathbf{T}_c(0, a_2)$ (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 ν_1, ν_2 which in turn implies $\mathcal{T}_c(\mu|\nu_2) = 0$ for all ν_2 which is impossible. Finally, using the convention 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 tensorization property. Moreover, if duality holds for the cost function c (in the sense of Definition 2.4), we can state a dual characterization of $\mathbf{T}_c(a_1, a_2)$ in the spirit of Bobkov-Götze dual formulation [8].

We now state these properties and characterizations.

4.2. Bobkov-Götze dual characterization.

The following characterization extends [8], thanks to the dual formulation of the transport cost; see also [24].

Proposition 4.5 (Dual formulation)**.** *Let $c\colon X \times \mathcal{P}_\gamma(X) \to [0, \infty]$ be a measurable cost function. Assume that $c(x, \delta_x) = 0$ for all $x \in X$ and that duality holds for the cost function c. For $\mu \in \mathcal{P}_\gamma(X)$ and $a_1, a_2, b > 0$, Items (i)'s and (ii)'s are equivalent:*

- *(i) $\mathbf{T}_c(a_1, a_2)$ holds;*

(ii) for all $\varphi \in \Phi_{\gamma,b}(X)$ (resp. for all non-negative $\varphi \in \Phi_\gamma$), it holds

$$\left(\int \exp\left\{\frac{R_c\varphi}{a_2}\right\} d\mu\right)^{a_2} \left(\int \exp\left\{-\frac{\varphi}{a_1}\right\} d\mu\right)^{a_1} \le 1; \tag{4.6}$$

- *(i′) $\mathbf{T}_c^+(b)$ holds;*
  (ii′) for all $\varphi \in \Phi_{\gamma,b}(X)$ (resp. for all non-negative $\varphi \in \Phi_\gamma$), it holds

$$\exp\left\{\int R_c\varphi\, d\mu\right\} \left(\int \exp\left\{\frac{-\varphi}{b}\right\} d\mu\right)^{b} \le 1; \tag{4.7}$$

- *(i″) $\mathbf{T}_c^-(b)$ holds;*
  (ii″) for all $\varphi \in \Phi_{\gamma,b}(X)$ (resp. for all non-negative $\varphi \in \Phi_\gamma$), it holds

$$\left(\int \exp\left\{\frac{R_c\varphi}{b}\right\} d\mu\right)^{b} \exp\left\{-\int \varphi\, d\mu\right\} \le 1\,, \tag{4.8}$$

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

Moreover, specializing to the "bar" cost $\overline{\mathcal{T}}_\theta$, 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 φ that are convex, Lipschitz and bounded from below.

Remark 4.9.

- *The preceding result thus applies to the cost functions defined in Section 2.4 under the assumptions of Theorems 2.7, 2.11 and 2.14 and more generally to all the cost functions satisfying the assumptions of our general duality in Theorem 9.6.*
- *In the result above, we implicitly assumed that functions $R_c\varphi$ were measurable. If it is not the case, then integrals of $R_c\varphi$ with respect to μ have to be replaced by integrals with respect to the inner measure μ_*.*
- *When $c(x,p) = \theta\left(x - \int y\, p(dy)\right)$, $x \in \mathbb{R}^m$, $p \in \mathcal{P}_1(\mathbb{R}^m)$, for some convex function $\theta : \mathbb{R}^m \to \mathbb{R}_+$, the inequality $\mathbf{T}_c(a_1,a_2)$ is thus equivalent to the following exponential type inequality first introduced by Maurey [44] (the so-called convex (τ)-property):*

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

Proof. By duality (*i.e.* using Definition 2.4), $\mathbf{T}_c(a_1,a_2)$ is equivalent to having

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

for all $\varphi \in \Phi_{\gamma,b}(X)$ and all $\nu_1, \nu_2 \in \mathcal{P}_\gamma(X)$. The expected result follows by taking the (two independent) suprema, on the left hand side, over ν_1 and ν_2, and by using Lemma 4.10 below. Note that since $c(x,\delta_x) = 0$ for all $x \in X$, one always has $R_c\varphi \le \varphi$, for all $\varphi \in \Phi_{\gamma,b}(X)$ and so the function $\psi = R_c\varphi/a_2$ satisfies the assumption of Lemma 4.10. This completes the proof of the equivalence $(i) \Leftrightarrow (ii)$.

Note that (4.6) is invariant under translations $\varphi \mapsto \varphi + a$ and so the functions φ 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 Theorem 2.11. □

Lemma 4.10. *Let $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function satisfying (2.1) and let $\mu \in \mathcal{P}_\gamma(X)$; for all measurable functions $\psi : \mathcal{X} \to \mathbb{R}$ such that $\psi \le \varphi$ for some $\varphi \in \Phi_\gamma(X)$, it holds*

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

This lemma can be interpreted as a particular case of Proposition 6.8 when $\mathcal{P}_\gamma(X) = \mathcal{P}(X)$ (and $\beta(x) = x\log x$, $x > 0$). For a sake of clarity we present below the simpler proof of this case.

Proof of Lemma 4.10. Consider the function $\beta(x) = x\log(x)$, $x > 0$. A simple calculation shows that $\beta^*(t) := \sup_{x>0}\{tx - \beta(x)\} = e^{t-1}$, $t \in \mathbb{R}$. Since $\psi \le \varphi$, 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 $xy \le \beta(x) + \beta^*(y)$, $x > 0, y \in \mathbb{R}$, one gets

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

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

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

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

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

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

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

when $k \to \infty$. This completes the proof. □

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

Theorem 4.11 (Tensoring property)**.** *Let $\gamma\colon \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function satisfying (2.1), $(X_1, d_1), \dots, (X_n, d_n)$ be complete separable metric spaces equipped with measurable cost functions $c_i : X_i \times \mathcal{P}_\gamma(X_i) \to [0, \infty]$, $i \in \{1, \dots, n\}$ 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 $\mu_i \in \mathcal{P}_\gamma(X_i)$ satisfy 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 $\mu_1 \otimes \cdots \otimes \mu_n$ 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 $c\colon X_1 \times \cdots \times X_n \times \mathcal{P}_\gamma(X_1 \times \cdots \times X_n) \to [0, \infty)$ defined by*

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

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

The following is an immediate corollary of Theorem 4.11.

Corollary 4.12. *Let $\gamma\colon \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous 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 $c : X \times \mathcal{P}_\gamma(X) \to [0, \infty]$ 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 $c^n : X^n \times \mathcal{P}_\gamma(X^n) \to [0, \infty)$ is the cost function defined by*

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

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

The proof of Theorem 4.11 is postponed to Appendix A.

5. Transport-entropy inequalities : link with dimension-free concentration

In this section, extending [23], 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 notation.

Let $\gamma \colon \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function satisfying (2.1) and $c : X \times \mathcal{P}_\gamma(X) \to [0, \infty)$ such that $c(x, \delta_x) = 0$ for all $x \in X$. Recall from Corollary 4.12 that for all integers $n \geq 1$,

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

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

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

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

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

and, for $t \geq 0$,

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

5.1. A general equivalence.

We are now in a position to state our theorem.

Theorem 5.1. *Let $\gamma \colon \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function satisfying (2.1) and $c : X \times \mathcal{P}_\gamma(X) \to [0, \infty)$ a measurable cost function such that $c(x, \delta_x) = 0$ for all $x \in X$, and for which duality holds in the sense of Definition 2.4. For $\mu \in \mathcal{P}_\gamma(X)$ and $a_1, a_2 > 0$, the following are equivalent:*

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

$$\mu^n(X^n \setminus A_t^n)^{a_2} \, \mu^n(A)^{a_1} \leq K e^{-t} \qquad \forall t \geq 0. \tag{5.2}$$

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

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

Moreover (i) $\Rightarrow$ (ii) with $K = 1$, (ii) $\Rightarrow$ (iii) with $K' = K$.

Remark 5.3.

- *The implication (i) $\Rightarrow$ (ii) was first discovered by Marton [41, 43, 42]. This nice observation is at the origin of the interest in transport-entropy inequalities.*
- *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 $p \mapsto c(x,p)$ is convex, as the proof indicates.*
- *Observe that the equivalence between (i), (ii) and (iii) implies that (iii) with any $K' \geq 1$ is equivalent to (iii) with $K' = 1$, and similarly (ii) with any $K \geq 1$ is equivalent to (ii) with $K = 1$.*

Proof. First we prove that (i) implies (ii). Since μ satisfies $\mathbf{T}_c(a_1, a_2)$, by the tensorization 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 ν_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

$$t \leq \mathcal{T}_{c^n}(\nu_1|\nu_2) \leq -a_1 \log(\mu^n(A)) - a_2 \log \mu^n(X^n \setminus A_t^n),$$

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 [25]. Fix $\varepsilon \in (0, 1)$. Given $f \in \Phi_\gamma(X)$, non-negative, let $\varphi(x) = f(x_1) + f(x_2) + \cdots + f(x_n)$, $x \in X^n$. 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 μ^n,

$$\left(\int e^{\frac{R_c f}{(1+\varepsilon)a_2}}\, d\mu\right)^{a_2} \left(\int e^{-\frac{f}{(1-\varepsilon)a_1}}\, d\mu\right)^{a_1} = \left(\int e^{\frac{R_{c^n}\varphi}{(1+\varepsilon)a_2}}\, d\mu^n\right)^{\frac{a_2}{n}} \left(\int e^{-\frac{\varphi}{(1-\varepsilon)a_1}}\, d\mu^n\right)^{\frac{a_1}{n}}. \tag{5.4}$$

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

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

and therefore

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

This in turn implies, by simple algebra that

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

where in the last line we used that φ is a non-negative function.

Plugging this bound into (5.4) leads, in the limit $n \to \infty$, 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 ε to 0 gives $\mathbf{T}_c(a_1,a_2)$, thanks to Proposition 4.5. □

5.2. **Particular cases.** In this section we focus on concentration inequalities related to the usual Monge-Kantorovich transport-cost and to barycentric transport-costs.

5.2.1. *Usual costs.* Note that when $c(x,p) = \int \omega(x,y)\,p(dy)$, for some measurable $\omega : X \times X \to [0,\infty)$, 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

$$B_r^n = \left\{x \in (\mathbb{R}^m)^n;\sum_{i=1}^n \|x_i\|^r \leq 1\right\}, \tag{5.5}$$

it holds

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

Concentration of measure inequalities are usually stated for enlargements of sets of measure bigger than 1/2 as in (1.2) (see [37]). In what follows we connect (5.2) to the usual definition for some families of cost functionals.

Lemma 5.6. *Consider a cost function c of the form*

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

with $\alpha : \mathbb{R}_+ \to \mathbb{R}_+$ an increasing convex function such that $\alpha(0) = \alpha'(0) = 0$ and satisfying the following doubling property : there exists $K \geq 2$ such that $\alpha(2x) \leq K\alpha(x)$ for all $x \in \mathbb{R}_+$. Suppose also that, for a given $n \in \mathbb{N}^*$, a probability measure μ on X satisfies, for some constants $a > 0, b \geq 1$, the following concentration property holds:

$$\mu^n(X^n \setminus A_t^n) \leq be^{-t/a}, \qquad \forall t \geq 0, \tag{5.7}$$

for all $A \subset X^n$ such that $\mu^n(A) \geq 1/2$.
Then μ satisfies the following property : for all $s \in (0,1)$ and for all $A \subset X^n$,

$$\mu^n(X^n \setminus A_t^n)^{1/(1-s)^{r-1}}\mu^n(A)^{1/s^{r-1}} \leq be^{-t/a}, \qquad \forall t \geq 0, \tag{5.8}$$

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

Conversely, if the concentration property (5.8) 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 $t > \max(a\log(2b),0)$,

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

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

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

To clarify the notations, we will omit some of the dependencies in n in this proof. The fact that (5.7) implies (5.8) is a consequence of the following set inclusions (that are justified at the end of the proof):

(a) $A \subset X^n \setminus ((X^n \setminus A_u)_u)$, $\forall u \geq 0$, and for all $s \in (0,1)$,

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

The last inclusion above follows from the identity,

$$\left(u^{1/r}+v^{1/r}\right)^r = \inf_{s\in(0,1)} \left\{ \frac{u}{s^{r-1}} + \frac{v}{(1-s)^{r-1}} \right\}. \tag{5.9}$$

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 (5.7), 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 (5.7), 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 (5.8) 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 $\sum_{i=1}^n \alpha(d(x_i,y_i)) \leq u$. But, since $y \in X^n \setminus A_u$, it holds $\sum_{i=1}^n \alpha(d(y_i,z_i)) > u$ for all $z \in A$. In particular, taking $z = x$, one gets a contradiction.
Finally, let us show (b). According to e.g. [27, Lemma 4.7], the function $x \mapsto \alpha^{1/r}(x)$ is subbadditive. It follows easily that $(x,y) \mapsto \left(\sum_{i=1}^n \alpha(d(x_i,y_i))\right)^{1/r}$ defines a distance on X^n. Point (b) then follows immediately from the triangle inequality. □

For the next corollary, recall the definition of B_r^n given in (5.5). Recall also that the median $\mathrm{med}(f)$ for a probability measure μ of a function f is any number (that need not be unique) such that $\mu(f \geq \mathrm{med}(f)), \mu(f \leq \mathrm{med}(f)) \geq 1/2$.

Corollary 5.10. *Let $r \geq 2$ and consider the cost $c(x,p) = \int \|x-y\|^r \, p(dy)$, $x \in \mathbb{R}^m$, $p \in \mathcal{P}_r(\mathbb{R}^m)$, where $\|\cdot\|$ is a norm on $\mathbb{R}^m$. For a probability measure $\mu \in \mathcal{P}_r(\mathbb{R}^m)$, the following assertions are equivalent :*

(1) There exist $a_1, b_1 > 0$ such that, $\forall n \in \mathbb{N}^*$,
$$\mu^n(A + t^{1/r}B_r^n) \geq 1 - b_1 e^{-t/a_1}, \qquad \forall t \geq 0,$$
for all sets A such that $\mu^n(A) \geq 1/2$.

(2) There exist $a_2, b_2 > 0$ such that, $\forall n \in \mathbb{N}^*$,
$$\mu^n(f > \mathrm{med}\,(f) + r) \leq b_2 e^{-t^r/a_2}, \qquad \forall t \geq 0,$$
for all $f : (\mathbb{R}^m)^n \to \mathbb{R}$ which are 1-Lipschitz with respect to the norm $\|\cdot\|_r^n$ defined on $(\mathbb{R}^m)^n$ by
$$\|x\|_r^n = \left(\sum_{i=1}^n \|x_i\|^r\right)^{1/r}, \qquad x \in (\mathbb{R}^m)^n.$$

(3) There exist $a_3, b_3 > 0$ such that, $\forall n \in \mathbb{N}^*$, $\forall s \in (0,1)$, and $\forall A \subset (\mathbb{R}^m)^n$,
$$\mu^n((\mathbb{R}^m)^n \setminus A_t^n)^{1/(1-s)^{r-1}} \mu^n(A)^{1/s^{r-1}} \leq b_3 e^{-t/a_3}, \qquad \forall t \geq 0,$$
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 $\forall s \in (0,1)$, μ satisfies
$$\mathbf{T}_r(a_4/s^{r-1}, a_4/(1-s)^{r-1}).$$

(5) $\exists a_5 > 0$ such that μ 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* $b_3 = b_1$, *and finally,* (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 [23, Theorem 1.3] for a similar statement involving more general cost functions.

Proof. The equivalence (1) $\Leftrightarrow$ (2) is very classical (see e.g [37, Proposition 1.3]).
The implications (1) $\Rightarrow$ (3) and (3) $\Rightarrow$ (1) are given in Lemma 5.6. (3) $\Rightarrow$ (4) and (4) $\Rightarrow$ (3) are consequences of Theorem 5.1.
If the property (4) holds, then for all $\nu_1 \in \mathcal{P}_r(\mathbb{R}^m)$,

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

As s goes to 1, we get (5), μ 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 $\nu_1, \nu_2 \in \mathcal{P}_r(\mathbb{R}^m)$,

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

The property (4) with $a_4 = a_5$ then follows from the identity (5.9). $\square$

5.2.2. *Barycentric costs.* 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

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

denoting by $\overline{\mathrm{conv}}(A)$ the closed convex hull of A and B_r^n as defined in (5.5). Indeed, denoting $\|\cdot\|_r^n$, for the norm defined on $(\mathbb{R}^m)^n$ by $\|x\|_r^n = \left(\sum_{i=1}^n \|x_i\|^r\right)^{1/r}$, then, for all $x \in (\mathbb{R}^m)^n$, it holds $c_A(x) = \inf_{y\in C}\{\|x-y\|_r^n\} = \inf_{y \in \overline{C}}\{\|x-y\|\}$, with $C = \{\int y\, p(dy); p \in \mathcal{P}_1(A)\}$. It is well known that $\overline{C} = \overline{\mathrm{conv}}\,(A)$, which proves the claim.

The result below (a consequence of Theorem 5.1) shows in particular that inequalities $\mathbf{T}_2^\pm$ are responsible for Gaussian dimension-free concentration for convex and concave Lipschitz functions.

Corollary 5.11. *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* $\mu \in \mathcal{P}_1(\mathbb{R}^m)$, *the following assertions are equivalent :*

(1) *There exist* $a_1, b_1 > 0$ *such that,* $\forall n \in \mathbb{N}^*$,
$$\mu^n(A + t^{1/r} B_r^n) \geq 1 - b_1 e^{-t/a_1}, \qquad \forall t \geq 0,$$
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,* $\forall n \in \mathbb{N}^*$,
$$\mu^n(f > \mathrm{med}\,(f) + t) \leq b_2 e^{-t^r/a_2}, \qquad \forall t \geq 0,$$
for all $f : (\mathbb{R}^m)^n \to \mathbb{R}$ *which is either* convex *or* concave *and* 1*-Lipschitz with respect to the norm* $\|\cdot\|_r^n$ *defined on* $(\mathbb{R}^m)^n$ *by*
$$\|x\|_r^n = \left(\sum_{i=1}^n \|x_i\|^r\right)^{1/r}, \qquad x \in (\mathbb{R}^m)^n.$$

(3) *There exist* $a_3, b_3 > 0$ *such that,* $\forall n \in \mathbb{N}^*$, $\forall s \in (0,1)$, *and* $\forall A \subset (\mathbb{R}^m)^n$,
$$\mu^n((\mathbb{R}^m)^n \setminus A_t^n)^{1/(1-s)^{r-1}} \mu^n(A)^{1/s^{r-1}} \leq b_3 e^{-t/a_3}, \qquad \forall t \geq 0,$$
where $A_t^n = \{x \in (\mathbb{R}^m)^n\,; c_A^n(x) \leq t\} = \overline{\mathrm{conv}}\, A + t^{1/r} B_r^n$.

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

(5) *There exists* $a_5 > 0$ *such that* μ *satisfies* $\overline{\mathbf{T}}_r^+(a_5)$ *and* μ *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 [37, Proposition 1.3], one sees easily that (1) $\Leftrightarrow$ (2), and, according to Theorem 5.1, (3) $\Leftrightarrow$ (4).

Let us show that (3) implies (1). Let A be a convex subset. As in Lemma 5.6, 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 the complement of convex sets.

We adapt the proof of Lemma 5.6 to get (1) $\Rightarrow$ (3). The property (a) is replaced by the following, for all subsets A,

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

Since $A_u = \overline{\mathrm{conv}}\, A + u^{1/r} B_r^n$, this property (a') is a simple consequence of the set property (a) applied to the set $\overline{\mathrm{conv}}\, A$. For the same reason, the set property (b) still holds. Then following the proof of Lemma 5.6, by using (a') and (b), with 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 it's complement $(\mathbb{R}^m)^n \setminus B$, we get (1) $\Rightarrow$ (3) with $a_3 = a_1$ and $b_3 = b_1$.

If the assertion (4) holds, then for all $\nu_1, \nu_2 \in \mathcal{P}_1(\mathbb{R}^m)$ and for all $\forall s \in (0,1)$,

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

As s goes to 1 or to 0, we get (5) – the fact that μ 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 $\nu_1, \nu_2 \in \mathcal{P}_1(\mathbb{R}^m)$,

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

□

6. Universal transport cost inequalities with respect to Hamming distance and Talagrand's concentration of measure inequalities

This section is devoted to universal transport-entropy inequalities associated to the weak transport costs $\widetilde{\mathcal{T}}$ and $\widehat{\mathcal{T}}$ with respect to the Hamming distance. These examples of transport inequalities were behind the present work, whose goal was also to unify different kinds of concentration results.

6.1. Transport inequalities for Marton's costs.

In this section, we recall a transport-entropy inequality obtained by Dembo [17], improving upon preceding works by Marton [42], and used in its dual form by the third-named author to reach concentration bounds for supremum of empirical processes around their mean with optimal constants in the deviation bounds (see [55]).

Let us introduce some notation. For $t \in (0,1)$, define α_t by

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

and also set $\alpha_0(u) = (1-u)\log(1-u) + u$ and $\alpha_1(u) = -u - \log(1-u)$ when $u \in (0,1)$ (and $+\infty$ otherwise). Let us consider the cost of the form $\widetilde{\mathcal{T}}$ associated to α_t:

$$\widetilde{\mathcal{T}}_{\alpha_t}(\nu_1|\nu_2) = \inf \int \alpha_t \left(\int \mathbf{1}_{x\neq y}\, p_x(dy) \right) \nu_2(dx)\,,$$

where the infimum runs over the set of kernels p such that $\nu_2 p = \nu_1$.

Theorem 6.1. *Let (X,d) be a Polish space, $t \in (0,1)$ and $\mu \in \mathcal{P}(X)$. Then, for all $\nu_1, \nu_2 \in \mathcal{P}(X)$, it holds*

$$\widetilde{\mathcal{T}}_{\alpha_t}(\nu_1|\nu_2) \leq \frac{1}{1-t} H(\nu_1|\mu) + \frac{1}{t} H(\nu_2|\mu). \tag{6.2}$$

For $t=0$, it also holds

$$\widetilde{\mathcal{T}}_{\alpha_0}(\nu_1|\mu) \leq H(\nu_1|\mu),$$

and for $t=1$,

$$\widetilde{\mathcal{T}}_{\alpha_1}(\mu|\nu_2) \leq H(\nu_2|\mu).$$

The transport inequality (6.2) is due to Dembo [17, Theorem 1.(i)]. A short proof of this theorem is given in [55] (see Lemma 2.1.) As shown in [55], the behavior of the family of cost functions α_t allows to capture optimal bounds for the deviations of suprema of empirical bounded processes.

Let us just recall simple and useful corollaries of Theorem 6.1. First observing that $\alpha_t(u) \geq u^2/2$ (since $\alpha_t''(u) \geq 1$), we immediately recover using Theorem 5.1 (implication (i) $\Rightarrow$ (ii)) the following celebrated concentration result by Talagrand (see [62, Theorem 4.1.1]).

Corollary 6.3. *For any probability measure μ on X, it holds*

$$\mu^n(X^n \setminus A_t^n) \leq \frac{1}{\mu^n(A)^{s/(1-s)}} e^{-st/2}, \qquad \forall t > 0, \forall s \in (0,1),$$

for all $A \subset X^n$ and $n \in \mathbb{N}^*$, where

$$A_t^n = \left\{ y \in X^n : \exists p \in \mathcal{P}(X^n) \text{ with } p(A) = 1 \text{ such that } \sum_{i=1}^n \left(\int \mathbf{1}_{x_i \neq y_i}\, p(dx) \right)^2 \leq t \right\}.$$

We refer to [62, 37, 2, 20, 51] for applications of this concentration inequality, under the so-called convex hull distance.

Corollary 6.4. *Suppose that μ is a probability on $\mathbb{R}^m$ (equipped with an arbitrary norm $\|\cdot\|$) such that the diameter of $\operatorname{supp}(\mu)$ is bounded by $M > 0$. Then μ satisfies the inequality $\widetilde{\mathbf{T}}_2(4M^2, 4M^2)$ and thus $\overline{\mathbf{T}}_2(4M^2, 4M^2)$ given in Definition 4.1.*

Proof. Observe that $\alpha_t(u) \geq u^2/2$, for all $u \in [0,1]$ and $t = 1/2$. Furthermore, if ν_1, ν_2 are absolutely continuous with respect to μ then $\operatorname{supp}(\nu_i) \subset \operatorname{supp}(\mu)$. Therefore, if $\pi(dxdy) = \nu_1(dx)p_x(dy)$ is a coupling between ν_1 and ν_2, then $\int \|x-y\|\, p_x(dy) \leq M \int \mathbf{1}_{\{x\neq y\}}\, p_x(dy)$, for ν_1-almost all x, and so

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

Optimizing over all π, and then using Theorem 6.1 for $t = 1/2$, completes the proof. □

We recover from the preceding result, and Corollary 5.11 (that guarantees from the inequality $\overline{\mathbf{T}}_2(4M^2, 4M^2)$ a concentration result), the well-known fact that any probability measure with a bounded support satisfies dimension-free Gaussian type concentration for convex/concave Lipschitz functions.

6.2. **Transport inequalities for Samson's costs.** Now we consider a stronger variant of Theorem 6.1 involving costs of the form $\widehat{\mathcal{T}}$. To state this result we need to introduce additional notation. For $t \in (0,1)$, one sets

$$\beta_t(u) := \sup_{s\in\mathbb{R}} \{su - \beta_t^*(s)\}, \qquad u \in \mathbb{R}.$$

where β_t^* is defined by

$$\beta_t^*(s) := \frac{te^{(1-t)s} + (1-t)e^{-ts} - 1}{t(1-t)}, \qquad s \in \mathbb{R}.$$

The function β_t is by definition the Legendre transform of β_t^*. Since β_t^* is convex and lower semi-continuous, β_t^* is also the Legendre transform of β_t (see Theorem 26.5 [**?**] (Fenchel-Moreau Theorem)). We extend the definition for $t \in \{0,1\}$ by setting

$$\beta_0^*(s) = e^s - s - 1 \qquad \text{and} \qquad \beta_1^*(s) = e^{-s} + s - 1, \qquad \forall s \in \mathbb{R}.$$

In general, β_t does not have an explicit expression, but for $t \in \{0,1\}$ an easy calculation shows that

$$\beta_0(u) = (1+u)\log(1+u) - u, \quad u \geq -1$$
$$\beta_1(u) = \beta_0(-u) = (1-u)\log(1-u) + u, \quad u \leq 1.$$

Finally, consider the cost of the form $\widehat{\mathcal{T}}$ associated to these functions:

$$\widehat{\mathcal{T}}_{\beta_t}(\nu_1|\nu_2) = \inf \iint \beta_t\left(\mathbf{1}_{x\neq y}\frac{dp_x}{d\mu}(dy)\right)\mu(dy)\nu_2(dx),$$

where the infimum runs over the set of kernels p such that $\nu_2 p = \nu_1$, with in addition, $p_x \ll \mu$ on $X \setminus \{x\}$, for ν_2-almost all $x \in X$.

Theorem 6.5. *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, for all $\nu_1, \nu_2 \in \mathcal{P}(X)$, it holds*

$$\widehat{\mathcal{T}}_{\beta_t}(\nu_1|\nu_2) \leq \frac{1}{1-t}H(\nu_1|\mu) + \frac{1}{t}H(\nu_2|\mu). \tag{6.6}$$

For $t=0$, it also holds

$$\widehat{\mathcal{T}}_{\beta_0}(\nu_1|\mu) \leq H(\nu_1|\mu),$$

and for $t=1$,

$$\widehat{\mathcal{T}}_{\beta_1}(\mu|\nu_2) \leq H(\nu_2|\mu).$$

By Proposition 4.5 and Theorem 2.14, one sees that Theorem 6.5 is exactly the dual form of Theorem 1.1 of [55] (for $n=1$). This new expected formulation of Theorem 1.1 in [55] is therefore a direct consequence of the generalization of the Kantorovich theorem (Theorem 9.6).

A direct consequence of Theorem 6.5 and implication (i) $\Rightarrow$ (ii) of Theorem 5.1 is the following deep concentration result that improves the one by Talagrand [63, Theorem 4.2.].

Corollary 6.7. *For any probability measure μ on X, it holds*

$$\mu^n(X^n \setminus A^n_{s,t}) \leq \frac{1}{\mu^n(A)^{s/(1-s)}}e^{-st}, \qquad \forall t>0, \quad \forall s \in (0,1),$$

for all $A \subset X^n$ and $n \in \mathbb{N}^*$, where

$$A^n_{s,t} = \Bigg\{ y \in X^n : \exists p \in \mathcal{P}(X^n) \text{ with } p(A)=1 \text{ and } p_i \ll \mu, \forall i$$
$$\text{such that } \sum_{i=1}^n \int \beta_s\left(\mathbf{1}_{x_i\neq y_i}\frac{dp_i}{d\mu}(x_i)\right)\mu(dx) \leq t \Bigg\},$$

with p_i denoting the i-th marginal of p.

In Talagrand's paper [63], this kind of concentration result is the main ingredient to get deviation inequalities of Bernstein-type for suprema of centered bounded empirical processes. Starting from the optimal transport inequality of Theorem 6.5, the third-named author has obtained optimal constants in the Bernstein bounds for the deviations under and above the mean [55]. This transportation method is an alternative of the entropy method introduced by Ledoux [36], and then developed by many authors. We refer to the book by Boucheron, Lugosi and Massart [16] for more development in this field.

Below, we sketch the proof of Theorem 6.5, by revisiting and to some extent simplifying some of the arguments given in [55] with the help of the duality results developed in the present paper and in [25]. The first of these duality formulas is Kantorovich duality for the cost $\widehat{\mathcal{T}}$ given in Theorem 2.14. The second formula is more classical and is recalled in the following proposition.

Proposition 6.8. *Let $\beta : [0,\infty) \to \mathbb{R}$ be a lower-semicontinuous strictly convex and super-linear function (i.e. $\beta(x)/x \to +\infty$ as $x \to \infty$). Let μ be a probability measure on a Polish space X and denote by U_β the function defined on $\mathcal{P}(X)$ by*

$$U_\beta(\nu) = \int \beta\left(\frac{d\nu}{d\mu}\right) d\mu,$$

if ν is absolutely continuous with respect to μ and $+\infty$ otherwise. Then, for any bounded continuous function φ on X, it holds

$$\sup_{\nu\in\mathcal{P}(X)} \left\{ \int \varphi(x)\,\nu(dx) - U_\beta(\nu) \right\} = \inf_{t\in\mathbb{R}} \left\{ \int \beta^{\circledast}(\varphi(x)+t)\,\mu(dx) - t \right\},$$

where $\beta^{\circledast}$ denotes the monotone conjugate *of β, defined by $\beta^{\circledast}(x) = \sup_{y\geq 0}\{xy - \beta(y)\}$, $x \in \mathbb{R}$.*

We refer to [25, Proposition 2.9] for a short and elementary proof of this result.

We begin with an elementary lemma connecting monotone and usual conjugates of our functions β_t. The proof is left to the reader.

Lemma 6.9. *For all $u \in \mathbb{R}$, $\beta_t^{\circledast}(u) = \beta_t^*([u]_+)$, where $[u]_+ = \max(0,u)$ is the positive part of u.*

The next lemma gives an expression of $\widehat{Q}_{\beta_t}\varphi$, which will be crucial in order to establish the dual form of the transport inequality.

Lemma 6.10. *Let (X,d) be a compact metric space or a countable set of isolated points and $t \in [0,1]$. For all bounded continuous functions $\varphi : X \to \mathbb{R}$, there exists a function $v : X \to \mathbb{R}$ such that $v(x) \leq \varphi(x)$ for all $x \in X$ and such that*

$$\widehat{Q}_{\beta_t}\varphi(x) = v(x) - \int \beta_t^*\left([v(x)-v(y)]_+\right)\mu(dy).$$

Proof. Fix $x \in X$ and recall that

$$\widehat{Q}_{\beta_t}\varphi(x) = \inf\left\{ \int \varphi(y)\,p(dy) + \int \beta_t\left(\mathbf{1}_{x\neq y}\frac{dp}{d\mu}(y)\right)\mu(dy) \right\},$$

where the infimum runs over the set of probability measures $p \ll \mu$ on $X \setminus \{x\}$.

A probability p of this set can be written $p = \alpha\delta_x + (1-\alpha)q$, with $\alpha = p(\{x\})$ and where q is another probability such that $q \ll \mu$ and $q(\{x\}) = 0$. So

$$\begin{aligned}(6.11)\quad &\widehat{Q}_{\beta_t}\varphi(x) - \varphi(x) \\ &= \inf_{\alpha\in[0,1]} \inf_{q\ll\mu, q(\{x\})=0} \left\{ \int (1-\alpha)(\varphi(y)-\varphi(x))\,q(dy) + \int \beta_t\left((1-\alpha)\mathbf{1}_{x\neq y}\frac{dq}{d\mu}(y)\right)\mu(dy) \right\}.\end{aligned}$$

Consider the probability measure μ_x with the following density with respect to μ: $\frac{d\mu_x}{d\mu}(y) = \lambda^{-1}\mathbf{1}_{x\neq y}$, where $\lambda = \mu(X\setminus\{x\}) > 0$ (we assume of course that μ is not the Dirac mass at point

x), then $q(\{x\}) = 0$ and $q \ll \mu$ if and only if $q \ll \mu_x$ and in this case $\frac{dq}{d\mu_x} = \lambda \frac{dq}{d\mu}$, μ_x-almost everywhere. Therefore, (6.11) becomes

$$\widehat{Q}_{\beta_t}\varphi(x) - \varphi(x) = \inf_{\alpha\in[0,1]} \inf_{q\ll\mu_x} \left\{ \int (1-\alpha)(\varphi(y)-\varphi(x))\, q(dy) + \lambda \int \beta_t \left(\frac{(1-\alpha)}{\lambda} \frac{dq}{d\mu_x}(y) \right) \mu_x(dy) \right\}.$$

So it holds

$$\begin{aligned}
\widehat{Q}_{\beta_t}\varphi(x) - \varphi(x) &= \inf_{\alpha\in[0,1)} \inf_{q\ll\mu_x} \left\{ \int (1-\alpha)(\varphi(y)-\varphi(x))\, q(dy) + \lambda \int \beta_t \left(\frac{(1-\alpha)}{\lambda} \frac{dq}{d\mu_x}(y) \right) \mu_x(dy) \right\} \\
&= \inf_{\alpha\in[0,1)} - \inf_{r\in\mathbb{R}} \left\{ \lambda \int \beta_t^{\circledast} \left(\frac{(1-\alpha)(\varphi(x)-\varphi(y)) + r}{(1-\alpha)} \right) \mu_x(dy) - r \right\} \\
&= \inf_{\alpha\in[0,1)} - \inf_{v\in\mathbb{R}} \left\{ \int \beta_t^* \left([v-\varphi(y)]_+ \right) \mathbf{1}_{x\neq y}\, \mu(dy) - (1-\alpha)(v-\varphi(x)) \right\} \\
&= \inf_{\alpha\in[0,1]} \sup_{v\in\mathbb{R}} \left\{ (1-\alpha)(v-\varphi(x)) - \int \beta_t^* \left([v-\varphi(y)]_+ \right) \mathbf{1}_{x\neq y}\, \mu(dy) \right\} \\
&= \sup_{v\in\mathbb{R}} \inf_{\alpha\in[0,1]} \left\{ (1-\alpha)(v-\varphi(x)) - \int \beta_t^* \left([v-\varphi(y)]_+ \right) \mathbf{1}_{x\neq y}\, \mu(dy) \right\} \\
&= \sup_{v\in\mathbb{R}} \left\{ -[v-\varphi(x)]_- - \int \beta_t^* \left([v-\varphi(y)]_+ \right) \mathbf{1}_{x\neq y}\, \mu(dy) \right\},
\end{aligned}$$

where the second equality comes from Proposition 6.8 and Lemma 6.9, and the second to the last one from (an elementary version of) the Min-Max theorem (see e.g., Corollary 37.3.2 [**?**]). In particular,

$$\begin{aligned}
\widehat{Q}_{\beta_t}\varphi(x) &= \varphi(x) - [v(x)-\varphi(x)]_- - \int \beta_t^* \left([v(x)-\varphi(y)]_+ \right) \mathbf{1}_{x\neq y}\, \mu(dy), \\
&= \min(v(x),\varphi(x)) - \int \beta_t^* \left([v(x)-\varphi(y)]_+ \right) \mathbf{1}_{x\neq y}\, \mu(dy).
\end{aligned}$$

for some function v (realizing the supremum in the last identity).

For a fixed $x \in X$, consider the function $F(v) = -[v-\varphi(x)]_- - \int \beta_t^* ([v-\varphi(y)]_+) \mathbf{1}_{x\neq y}\, \mu(dy)$, $v \in \mathbb{R}$. Since β_t^* is increasing on $[0,\infty)$, the function F is clearly non-increasing on $[\varphi(x), +\infty)$. Therefore F reaches its supremum on $(-\infty, \varphi(x)]$. On $(-\infty, \varphi(x))$, the function F is differentiable and it holds

$$\begin{aligned}
F'(v) &= 1 - \int e^{(1-t)[v(x)-\varphi(y)]_+} \mathbf{1}_{v(x)>\varphi(y)} \mathbf{1}_{x\neq y}\, \mu(dy) + \int e^{-t[v(x)-\varphi(y)]_+} \mathbf{1}_{v(x)>\varphi(y)} \mathbf{1}_{x\neq y}\, \mu(dy) \\
&= 1 - \int e^{(1-t)[v(x)-\varphi(y)]_+} \mathbf{1}_{x\neq y}\, \mu(dy) + \int e^{-t[v(x)-\varphi(y)]_+} \mathbf{1}_{x\neq y}\, \mu(dy).
\end{aligned}$$

It is not difficult to prove the existence of a point $\bar{v}$ (independent of x) such that

$$\int e^{(1-t)[\bar{v}-\varphi(y)]_+} \mathbf{1}_{x\neq y}\, \mu(dy) = 1 + \int e^{-t[\bar{v}-\varphi(y)]_+} \mathbf{1}_{x\neq y}\, \mu(dy) \tag{6.12}$$

and to check that the function F reaches its supremum at $v(x) := \min(\bar{v}, \varphi(x))$.

Finally, note that $[v(x)-\varphi(y)]_+ = [v(x)-v(y)]_+$, which completes the proof. □

The next result is Lemma 2.2 of [55].

Lemma 6.13. *Let μ be a probability measure on a measurable space X. For every bounded function $v : X \to \mathbb{R}$, it holds for all $t \in [0,1]$,*

$$\left(\int e^{tv(x) - t\int \beta_t^*([v(x)-v(y)]_+)\mu(dy)} \mu(dx) \right)^{1/t} \left(\int e^{-(1-t)v(x)} \mu(dx) \right)^{1/(1-t)} \leq 1.$$

With these lemmas in hand, we are now in a position to prove Theorem 6.5.

Proof Theorem 6.5. Fix $t \in (0,1)$; according to Proposition 4.5, the transport inequality (6.6) is equivalent to proving that

$$\left(\int e^{t\widehat{Q}_{\beta_t}\varphi(x)}\,\mu(dx)\right)^{1/t}\left(\int e^{-(1-t)\varphi(x)}\,\mu(dx)\right)^{1/(1-t)} \le 1, \tag{6.14}$$

for all bounded continuous functions $\varphi : X \to \mathbb{R}$. But according to Lemma 6.10,

$$\widehat{Q}_{\beta_t}\varphi(x) = v(x) - \int \beta_t^*([v(x)-v(y)]_+)\,\mu(dy),$$

for some function $v \le \varphi$ (possibly depending on t and on μ). According to Lemma 6.13, it holds

$$\left(\int e^{t\widehat{Q}_{\beta_t}\varphi(x)}\,\mu(dx)\right)^{1/t}\left(\int e^{-(1-t)v(x)}\,\mu(dx)\right)^{1/(1-t)} \le 1.$$

Since $v \le \varphi$, this gives (6.14) and completes the proof. □

Now for the sake of completeness, we give a quick proof of Lemma 6.13 in the particular case $t = 1$. The general case is more tricky and the interested reader is referred to [55].

Proof of Lemma 6.13 for $t=1$. In this case, the conclusion of the lemma amounts to proving that for all bounded measurable $v : X \to \mathbb{R}$, it holds

$$\int e^{H(v(x))}d\mu(x) \le 1, \tag{6.15}$$

where

$$H(v(x)) = v(x) - \int v(y)d\mu(y) - D(v(x)),$$

with $D(v(x)) = \int \beta_1^*([v(x)-v(y)]_+)d\mu(y)$. Replacing everywhere v by λv, $\lambda \ge 0$, it is equivalent to showing that for all $\lambda \ge 0$,

$$\phi(\lambda) = \int e^{H(\lambda v(x))}d\mu(x) \le 1.$$

Since $\phi(0) = 1$, it is sufficient to get that $\phi'(\lambda) \le 0$ for all $\lambda \ge 0$. Let us first observe that since $\beta_1^*(h) = e^{-h} + h - 1$,

$$H(\lambda v(x)) = \int \left(1 - e^{-\lambda[v(x)-v(y)]_+}\right)d\mu(y) - \int \lambda[v(y)-v(x)]_+ d\mu(y).$$

It follows that for $\lambda \ge 0$,

$$\begin{aligned}\phi'(\lambda) &= \int\left(\int [v(x)-v(y)]_+ e^{-\lambda[v(x)-v(y)]_+}d\mu(y) - \int [v(y)-v(x)]_+ d\mu(y)\right)e^{H(\lambda v(x))}d\mu(x)\\ &= \iint [v(x)-v(y)]_+\left(e^{-\lambda[v(x)-v(y)]_+ + H(\lambda v(x))} - e^{H(\lambda v(y))}\right)d\mu(x)d\mu(y)\,.\end{aligned}$$

For $v(x) \ge v(y)$ one has

$$-\lambda[v(x)-v(y)]_+ + H(\lambda v(x)) - H(\lambda v(y)) = D(\lambda v(y)) - D(\lambda v(x)) \le 0,$$

and therefore $\phi'(\lambda) \le 0$ for $\lambda \ge 0$. This ends the proof of (6.15). □

7. Discrete examples : Bernoulli, Binomial and Poisson laws

In this section, we gather some basic 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.

We first consider some results for the Bernoulli measure, derived in [54], and as such introduce some notations from there. Given $\rho \in (0,1)$, let $\mu_\rho := (1-\rho)\delta_0 + \rho\delta_1$ denote the non symmetric Bernoulli measure, and define $u_{\rho,0} : \mathbb{R} \to \mathbb{R}_+ \cup \{+\infty\}$ as

$$u_{\rho,0}(h) := \begin{cases} \frac{1-\rho(1-h)}{\rho}\log\frac{1-\rho(1-h)}{1-\rho} + (1-h)\log(1-h)\,, & \text{if } -\frac{1-\rho}{\rho} \le h \le 1\\ +\infty & \text{otherwise}\end{cases}$$

and define $u_{\rho,1}\colon \mathbb{R}\to\mathbb{R}_+\cup\{+\infty\}$ as

$$u_{\rho,1}(h) := \begin{cases} \frac{1}{\rho}\left[(1-\rho-h)\log\frac{1-\rho-h}{1-\rho} - (1-h)\log(1-h)\right], & \text{if } h\le 1-\rho \\ +\infty & \text{otherwise} \end{cases}$$

Finally we define, for $t\in\{0,1\}$,

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

One may simply check that $\theta_{\rho,t}$ is a lower semi-continuous convex function on $\mathbb{R}$ for any $t\in\{0,1\}$.

In [54], a family of cost functions $(\theta_{\rho,t},\, t\in[0,1])$ is introduced, that interpolates between $\theta_{\rho,0}$ and $\theta_{\rho,1}$ (see Section 2.4). According to Lemma 2.3 in [54], one has for any convex function $\varphi:\mathbb{R}\to\mathbb{R}$, and for any $t\in(0,1)$,

$$\left(\int e^{tQ_{\theta_{\rho,t}}\varphi}\,d\mu\right)^{1/t}\left(\int e^{-(1-t)\varphi}\,d\mu\right)^{1/(1-t)} \le 1, \tag{7.1}$$

where $Q_{\theta_{\rho,t}}\varphi$ is the infimum-convolution operator associated to the cost function $\theta_{\rho,t}$. The function $\theta_{\rho,t}$ is constructed as an optimal choice in (7.1). It can be shown that $\theta_{\rho,t}$ is convex and lower semi-continuous. As a consequence, according to the last point of Remark 4.9, the property (7.1) is equivalent to a family of weak transport inequalities with transport cost $\overline{\mathcal{T}}_{\theta_{\rho,t}}$ for $t\in(0,1)$, as in Theorems 6.1 and 6.5. To simplify, we only consider below the limiting cases for $t=0$ and $t=1$.

Proposition 7.2. *For all $\rho\in(0,1)$, it holds*

$$\overline{\mathcal{T}}_{\theta_{\rho,1}}(\mu_\rho|\nu)\le H(\nu|\mu_\rho) \quad \textit{and} \quad \overline{\mathcal{T}}_{\theta_{\rho,0}}(\nu|\mu_\rho)\le H(\nu|\mu_\rho), \quad \forall \nu\in\mathcal{P}(\{0,1\}). \tag{7.3}$$

By Theorem 4.11, the $\overline{\mathcal{T}}$-transport-entropy inequalities for the Bernoulli measure μ_ρ given in this proposition tensorize. Hence, the product of Bernoulli measures $\mu_\rho^n := \mu_\rho\otimes\cdots\otimes\mu_\rho$ on the hypercube $\{0,1\}^n$ satisfies the following n-dimensional version of the $\overline{\mathcal{T}}$-transport-entropy inequalities. First, recall that the corresponding n-dimensional costs are defined, for all $x=(x_1,\ldots,x_n)\in\{0,1\}^n$ and all $p\in\mathcal{P}(\{0,1\}^n)$, respectively by

$$\bar{c}^{(n)}_{\rho,t}(x,q) := \sum_{i=1}^n \theta_{\rho,t}\left(x_i - \int_{\{0,1\}} y_i\, q_i(dy_i)\right).$$

where $q_i\in\mathcal{P}(\{0,1\})$ is the i-th marginal of q, and $t\in\{0,1\}$. We denote by $\overline{\mathcal{T}}_{\bar{c}^{(n)}_{\rho,t}}$ the corresponding transport cost. Applying Theorem 4.11, we immediately get, from Proposition 7.2, the following $\overline{\mathcal{T}}$-transport-entropy inequalities for product of Bernoulli measures.

Corollary 7.4. *For all $\rho\in(0,1)$ and all integers $n\ge1$, it holds*

$$\overline{\mathcal{T}}_{\bar{c}^{(n)}_{\rho,1}}(\mu_\rho^n|\nu)\le H(\nu|\mu_\rho^n) \quad \textit{and} \quad \overline{\mathcal{T}}_{\bar{c}^{(n)}_{\rho,0}}(\nu|\mu_\rho^n)\le H(\nu|\mu_\rho^n), \quad \forall\nu\in\mathcal{P}(\{0,1\}^n).$$

Projection arguments are useful tools to reach transport-entropy inequalities as explained in the seminal work by Maurey (see Lemma 2 [44]). Indeed, that is the way we get $\overline{\mathcal{T}}$-transport inequalities for the binomial distribution $B(n,\rho)$, $\rho\in(0,1)$ from Corollary 7.4. 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 $k\in I_n := \{0,1,\ldots,n\}$. By using the fact that the image measure of μ_ρ^n by the projection $\phi:\{0,1\}^n\ni(x_1,\ldots,x_n)\mapsto\sum_{i=1}^n x_i\in I_n$ is the measure $\mu_{n,\rho}$, i. e. $\phi\#\mu_\rho^n=\mu_{n,\rho}$, Corollary 7.4 provides the following $\overline{\mathcal{T}}$-transport inequalities for $\mu_{n,\rho}$.

Corollary 7.5. *For all $\rho\in(0,1)$ and all integers $n\ge1$, it holds*

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

where $\theta_{\rho,t,n}(h) := n\theta_{\rho,t}(h/n)$, $h\in\mathbb{R}$, $t\in\{0,1\}$.

Proof. To simplify the notations, we write $\theta_{\rho,t} = \theta$, $\theta_{\rho,t,n} = \theta_n$, $\bar{c}^{(n)}_{\rho,t} = \bar{c}^{(n)}$, $\mu_\rho = \mu$ and $\mu_{n,\rho} = \mu_n$. The result is based on the following contraction argument : for any probability measure ν_2 and ν_1 on I_n with respective densities f_1 and f_2 with respect to μ_n, if $\nu_1 << \nu_2$ then one has

$$\overline{\mathcal{T}}_{\theta_n}(\nu_1|\nu_2) \le \overline{\mathcal{T}}_{\bar{c}^{(n)}}(\hat{\nu}_1|\hat{\nu}_2), \tag{7.6}$$

where $\hat{\nu}_1$ and $\hat{\nu}_2$ are the probability measures on $\{0,1\}^n$ with respective densities $f_1 \circ \phi$, and $f_2 \circ \phi$ with respect to μ^n. Observe that since $\mu_n = \phi\#\mu^n$, we also have $\nu_1 = \phi\#\hat{\nu}_1$ and $\nu_2 = \phi\#\hat{\nu}_2$.

Then, the first transport-entropy inequality of Corollary 7.5 follows from the contraction property (7.6) applied with $\nu_1 = \mu_n$ and $\nu_2 = \nu$, and from the first inequality of Corollary 7.4 observing that

$$H(\nu|\mu_{n,\rho}) = H(\nu_2|\nu_1) = H(\hat{\nu}_2|\hat{\nu}_1).$$

We follow the same arguments to recover the second inequality of Corollary 7.5.

Let us now turn to the proof of (7.6). Given $\hat{\pi} \in \Pi(\hat{\nu}_2, \hat{\nu}_1)$, there exists a probability kernel $\hat{p}$ satisfying for all $x, y \in \{0,1\}^n$, $\hat{\pi}(x,y) = \hat{\nu}_2(x)\hat{p}_x(y)$. Let π denote the probability measure on $I_n \times I_n$ defined by

$$\pi(w,z) := \hat{\pi}(\phi^{-1}(\{w\}) \times \phi^{-1}(\{z\})), \qquad w, z \in I_n.$$

This measure π has marginals ν_2 and ν_1, and for any $w, z \in I_n$, $\pi(w,z) = \nu_2(w)p_w(z)$, with

$$p_w(z) = \frac{1}{\nu_2(w)} \sum_{x,\phi(x)=w} \sum_{y,\phi(y)=z} \hat{p}_x(y)\hat{\nu}_2(x),$$

if $\nu_2(w) \ne 0$. From the definition of $\overline{\mathcal{T}}_{\theta_n}(\nu_1|\nu_2)$ and by Jensen inequality, it follows that

$$\begin{aligned}
\overline{\mathcal{T}}_{\theta_n}(\nu_1|\nu_2) &\le \int \theta_n\left(w - \int z\, p_w(dz)\right)\nu_2(dw) \\
&= \sum_{w\in I_n} \theta_n\left(w - \sum_{z\in I_n}\sum_{x,\phi(x)=w}\sum_{y,\phi(y)=z} z\,\hat{p}_x(y)\,\frac{\hat{\nu}_2(x)}{\nu_2(w)}\right)\nu_2(w) \\
&= \sum_{w\in I_n} \theta_n\left(\sum_{x,\phi(x)=w}\sum_{y\in\{0,1\}^n} (\phi(x)-\phi(y))\,\hat{p}_x(y)\,\frac{\hat{\nu}_2(x)}{\nu_2(w)}\right)\nu_2(w) \\
&\le \sum_{w\in I_n}\sum_{x,\phi(x)=w} \theta_n\left(\sum_{y\in\{0,1\}^n} (\phi(x)-\phi(y))\,\hat{p}_x(y)\right)\hat{\nu}_2(x) \\
&= \int \theta_n\left(\int (\phi(x)-\phi(y))\,\hat{p}_x(dy)\right)\hat{\nu}_2(dx).
\end{aligned}$$

The proof of (7.6) ends since, by convexity of the function θ,

$$\theta_n\left(\int (\phi(x)-\phi(y))\,\hat{p}_x(dy)\right) \le \sum_{i=1}^n \theta\left(\int (x_i - y_i)\,\hat{p}_x(dy)\right) = \bar{c}^{(n)}(x,\hat{p}_x),$$

and then optimizing over all $\hat{\pi} \in \Pi(\hat{\nu}_2,\hat{\nu}_1)$ with $\hat{\pi}(x,y) = \hat{\nu}_2(x)\hat{p}_x(y)$. □

Let p_λ denote the Poisson probability measure with parameter $\lambda > 0$, $p_\lambda(k) := \frac{\lambda^k}{k!}e^{-\lambda}$, $k \in \mathbb{N}$. By the weak convergence of the binomial distribution μ_{n,ρ_n}, with $\rho_n := \lambda/n$, towards the Poisson measure p_λ, Corollary 7.5 implies the next $\overline{\mathcal{T}}$-transport inequalities for the measure p_λ.

Set, for $t \in \{0,1\}$, $h \in \mathbb{R}$, $c_{\lambda,t}(h) := \lim_{n\to\infty} n\theta_{\rho_n,t}\left(\frac{h}{n}\right)$. One has

$$c_{\lambda,0}(h) = \lambda\, w\left(\frac{h}{\lambda}\right)\mathbf{1}_{h\le 0}, \quad \textit{and} \quad c_{\lambda,1}(h) := \lambda\, w\left(\frac{-h}{\lambda}\right)\mathbf{1}_{h\le 0},$$

where $w(h) = (1-h)\log(1-h) + h$ for $h \le 1$ and $w(h) = +\infty$ if $h > 1$.

Proposition 7.7. *For all $\lambda > 0$, it holds*

$$\overline{\mathcal{T}}_{c_{\lambda,0}}(p_\lambda|\nu) \le H(\nu|p_\lambda), \quad \textit{and} \quad \overline{\mathcal{T}}_{c_{\lambda,1}}(\nu|p_\lambda) \le H(\nu|p_\lambda), \qquad \forall \nu \in \mathcal{P}_1(\mathbb{N}).$$

Remark 7.8.

- *Observe that given $\lambda > 0$ and $t \in \{0,1\}$, there exist $a > 0$ and $b \in \mathbb{R}$ such that for any $h \in \mathbb{R}$,*
$$a[-h]_+ + b \leq c_{\lambda,t}(h),$$
where $[h]_+ = \max(0,h)$. By Jensen inequality, it follows that condition (C_4) in Definition 9.3 holds, namely, if $\nu_2 \in \mathcal{P}_1(\mathbb{N})$ and $(p_\ell)_{\ell\in\mathbb{N}}$ are probability kernels such that $p_\ell \in \mathcal{P}_1(\mathbb{N})$ for all $\mathbb{N}$ and
$$\int c_{\lambda,t}\left(\ell - \int k\, p_\ell(dk)\right)\nu_2(d\ell) < \infty,$$
then, setting $\nu_1 = \nu_2 p$, one has
$$a\left[\int k\,\nu_1(dk) - \int \ell\,\nu_2(d\ell)\right]_+ + b \leq a\int\left[-\ell + \int k\,p_\ell(dk)\right]_+ \nu_2(d\ell) + b < \infty\,,$$
and therefore $\int k\,\nu_1(dk) < \infty$, $\nu_1 \in \mathcal{P}_1(\mathbb{N})$.
As a consequence, following the proof of Theorem 2.11, one may check that the dual formula still holds for the weak optimal cost $\overline{\mathcal{T}}_{c_{\lambda,t}}$, namely
$$\overline{\mathcal{T}}_{c_{\lambda,t}}(\nu_1|\nu_2) = \sup_{\varphi\in\Phi_{1,b}(\mathbb{N})}\left\{\int \overline{Q}_{c_{\lambda,t}}\varphi\,d\nu_2 - \int\varphi\,d\nu_1\right\},\quad \nu_1,\nu_2 \in \mathcal{P}_1(\mathbb{N}). \tag{7.9}$$
By monotone convergence, the supremum can be restricted to all bounded functions.
- *The $\overline{\mathcal{T}}$-transport inequalities of Proposition 7.7 are optimal, i.e. the constant 1 cannot be improved. Indeed, e.g. the second inequality Proposition 7.7 is equivalent, thanks to Proposition 4.5, to*
$$\exp\left\{\int \overline{Q}_{c_{\lambda,0}}\varphi\,dp_\lambda\right\}\int e^{-\varphi}\,dp_\lambda \leq 1\,,$$
for all $\varphi \in \Phi_{1,b}(\mathbb{N})$ which is an equality for $\varphi(x) = -tx$, $x \in \mathbb{R}$, $t \geq 0$ (the same holds for the first inequality).

Proof of Proposition 7.7. Let $\nu \in \mathcal{P}_1(\mathbb{N})$ with density f with respect to p_λ. For any integer $n \geq 1$, we define $\hat{\nu}^n \in \mathcal{P}(I_n)$ by
$$\hat{\nu}^n(k) := \frac{f(k)\mu_{n,\rho_n}(k)}{\sum_{l\in I_n} f(l)\mu_{n,\rho_n}(l)}.$$
By the weak convergence of μ_{n,ρ_n} towards p_λ, the sequence of measure $\hat{\nu}^n$ also weakly converges towards ν, and moreover, if $H(\nu|p_\lambda) < \infty$ then
$$\lim_{n\to\infty} H(\hat{\nu}^n|\mu_{n,\rho_n}) = H(\nu|p_\lambda).$$
As a consequence, Corollary 7.5 implies that
$$\liminf_{n\to\infty}\overline{\mathcal{T}}_{\theta_{\rho,1,n}}(\mu_{n,\rho}|\hat{\nu}^n) \leq H(\nu|p_\lambda), \quad \text{and} \quad \liminf_{n\to\infty}\overline{\mathcal{T}}_{\theta_{\rho,0,n}}(\hat{\nu}^n|\mu_{n,\rho}) \leq H(\nu|p_\lambda).$$
The proof of Proposition 7.7 ends by using the following property: if $\nu_1^n \in \mathcal{P}(I_n)$, respectively $\nu_2^n \in \mathcal{P}(I_n)$, are sequences of probability measures that converge weakly towards $\nu_1 \in \mathcal{P}_1(\mathbb{N})$, respectively $\nu_2 \in \mathcal{P}_1(\mathbb{N})$, then
$$\overline{\mathcal{T}}_{c_{\lambda,t}}(\nu_1|\nu_2) \leq \liminf_{n\to\infty}\overline{\mathcal{T}}_{\theta_{\rho,t,n}}(\hat{\nu}_1^n|\hat{\nu}_2^n), \qquad t \in \{0,1\}. \tag{7.10}$$

It remains to prove this inequality. Let $\varphi : \mathbb{N} \to \mathbb{R}$ be a bounded function and set $\|\varphi\|_\infty := \sup_{k\in\mathbb{N}}|\varphi(k)|$. The dual expression of $\overline{\mathcal{T}}_{\theta_{\rho,t,n}}(\hat{\nu}_1^n|\hat{\nu}_2^n)$ given by Theorem 2.11 implies
$$\overline{\mathcal{T}}_{\theta_{\rho,t,n}}(\hat{\nu}_1^n|\hat{\nu}_2^n) \geq \int \overline{Q}_{\theta_{\rho_n,t,n}}\varphi\,d\hat{\nu}_2^n - \int\varphi\,d\hat{\nu}_1^n. \tag{7.11}$$
Since $\theta_{\rho_n,t,n} \geq 0$ and $\theta_{\rho_n,t,n}(0) = 0$, it holds for $\ell \in I_n$
$$\overline{Q}_{\theta_{\rho_n,t,n}}\varphi(\ell) \geq \inf_{q\in\mathcal{P}(I_n)}\left\{\int_{I_n}\varphi\,dq + \theta_{\rho_n,t,n}\left(-\left[\ell - \int_{I_n} k\,q(dk)\right]_-\right)\right\},$$

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

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

At this point, we claim that for all $h \geq 0$ and $n \geq 2\lambda$,

$$\theta_{\rho_n,t,n}(-h) \geq v_t(h),$$

where $v_0(h) := \lambda w\left(-\frac{h}{2\lambda}\right)$ and $v_1(h) := \lambda w\left(-\frac{h}{\lambda}\right) - 2\lambda w\left(-\frac{h}{2\lambda}\right)$. This is an easy consequence of the fact that $\theta_{\rho_n,0,n}(-h) \geq \lambda w\left(-\frac{(1-\rho_n)h}{\lambda}\right)$ and $\theta_{\rho_n,0,n}(-h) \geq \lambda w\left(-\frac{h}{\lambda}\right) - nw\left(-\frac{h}{n}\right)$, for all $h \geq 0$. Let v_t^{-1} denote the inverse function of increasing bijection $v_t : \mathbb{R}_+ \to \mathbb{R}_+$. It follows that

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

In turn, since $\mathcal{P}(I_n) \subset \mathcal{P}_1(\mathbb{N})$,

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

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

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

As a consequence, we get from (7.11) that

$$\liminf_{n\to\infty} \overline{\mathcal{T}}_{\theta_{\rho,t,n}}(\hat{\nu}_1^n|\hat{\nu}_2^n) \geq \lim_{n\to\infty}\left[\int \overline{Q}_{c_{\lambda,t}}\varphi\, d\hat{\nu}_2^n - \int \varphi\, d\hat{\nu}_1^n\right] = \int \overline{Q}_{c_{\lambda,t}}\varphi\, d\nu_2 - \int \varphi\, d\nu_1,$$

where the last equality holds by weak convergence. The proof of (7.10) ends by optimizing over all bounded functions φ and from the identity (7.9). □

8. 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 4.1 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 8.1), we restrict to the specific case (already of interest) of $\overline{\mathcal{T}}_\theta$, introduced in Section 2.4. 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). As an application, using the characterization of $\overline{\mathbf{T}}_2^-$ by means of log-Sobolev type inequalities and results from [1], we may give more examples of measures satisfying such a transport-entropy inequality on the line.

8.1. Transport-entropy and (τ)-log-Sobolev inequalities.

In this section, we generalize the notion of (τ)-log-Sobolev inequality introduced in [26] (see also [27]) and describe some connection to weak transport-entropy inequalities.

First we need some notation. Given $\lambda > 0$ and $\varphi \in \Phi_\gamma(X)$, define

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

Observe that $R_c^1 = R_c$, where R_c is defined in Theorem 9.6. Following [26], we introduce the (τ)-log-Sobolev inequality as follows. We recall that for any non-negative function g, one denotes $\mathrm{Ent}_\mu(g) = \int g\log g\, d\mu - \left(\int g\, d\mu\right)\log\left(\int g\, d\mu\right)$.

Definition 8.1 ($(\tau) - \mathbf{LSI}_c(\lambda, C)$)**.** *Let $\gamma\colon \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function satisfying (2.1). Let $c : X \times \mathcal{P}_\gamma(X) \to [0,\infty)$ be a cost function and $C \in (0,\infty)$. Then $\mu \in \mathcal{P}_\gamma(X)$ is said to satisfy the (τ)-log-Sobolev inequality with constant C, λ and cost c (or in short $(\tau) - \mathbf{LSI}_c(\lambda, C)$) if, for all f with $\int f e^f d\mu < \infty$, it holds*

$$\mathrm{Ent}_\mu(e^f) \leq C \int (f - R_c^\lambda f) e^f \, d\mu. \tag{8.2}$$

The following result extends [26, Theorem 2.1].

Proposition 8.3. *Let $\gamma\colon \mathbb{R}_+ \to \mathbb{R}_+$ be continuous function satisfying (2.1) and $c : X \times \mathcal{P}_\gamma(X) \to [0,\infty)$ be a cost function. If $\mu \in \mathcal{P}_\gamma(X)$ satisfies $\mathbf{T}_c^-(b)$, then it satisfies $(\tau) - \mathbf{LSI}_c(\lambda, \frac{1}{1-\lambda b})$ for all $\lambda \in (0, 1/b)$.*

Remark 8.4. *In $\mathbb{R}^n$ [26], and more generally in metric spaces [27], if one considers the usual transport cost $\mathcal{T}_2$ (with cost $\omega(x,y) = d(x,y)^2$), it is proved that the corresponding $\mathbf{T}_2^-(b)$ is actually equivalent to some (τ)-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 [57] for some developments). This is the primary reason for us restricting ourselves to the specific case of the "bar" cost in the next sections.*

Proof. Fix a function $f : X \to \mathbb{R}$ 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 $\pi(dxdy) = \nu_f(dx) p_x(dy)$ is a probability measure on $X \times X$ with first marginal ν_f and second marginal μ,

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

It follows from the definition of R_c^λ 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,

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

Hence,

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

Optimizing over all π (or equivalently over all p_x) with marginals ν_f and μ, 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 proposition follows by noticing that $\left(\int e^f \, d\mu\right) H(\nu_f|\mu) = \mathrm{Ent}_\mu\left(e^f\right)$. □

8.2. **Weak transport-entropy inequalities $\overline{\mathcal{T}}_2^{\pm}$.** 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 [33] *restricted* to convex/concave functions, to the (τ)-log-Sobolev inequality (8.2) 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

$$c(x,p) = \frac{1}{2} \left\| x - \int y \, p(dy) \right\|^2, \qquad x \in \mathbb{R}^m, \ p \in \mathcal{P}_1(\mathbb{R}^m),$$

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 2.4 and the (τ)-log-Sobolev inequality

(8.2) defined with such a cost. As usual, $\|f\|_p := (\int |f|^p \, d\mu)^{\frac{1}{p}}$, $p \in \mathbb{R}^*$ (including negative real numbers) and $\|f\|_0 := \exp\{\int \log |f| \, d\mu\}$ whenever this makes sense. Also, given $\varphi \colon \mathbb{R}^m \to \mathbb{R}$, $t > 0$, let

$$Q_t\varphi(x) := \inf_{y \in \mathbb{R}^m} \left\{ \varphi(y) + \frac{1}{2t} \|x - y\|^2 \right\}, \quad x \in \mathbb{R}^m, \tag{8.5}$$

denote the Hamilton-Jacobi semi-group and

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

We will make use of the following observation (see Theorem 2.11): for any $\varphi \colon \mathbb{R}^m \to \mathbb{R}$ convex, Lipschitz and bounded from below, it holds

$$Q_1\varphi(x) = R_c\varphi(x) = \inf_{p \in \mathcal{P}_1(X)} \left\{ \int \varphi(y) \, p(dy) + c(x, p) \right\}, \quad x \in \mathbb{R}^m.$$

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

$$(x \neq y \text{ with } \|x\|_* = \|y\|_* = 1) \Rightarrow \|(1-t)x + ty\|_* < 1, \text{ for } 0 < t < 1. \tag{8.6}$$

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 8.12 below). The proof could certainly be adapted without this assumption, but we would rather not enter into these technical complications.

Remark 8.7. *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 8.8. *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 $\varphi \colon \mathbb{R}^m \to \mathbb{R}$ convex, Lipschitz and bounded from below, it holds

$$\mathrm{Ent}_\mu(e^\varphi) \leq \frac{1}{2\rho} \int \|\nabla\varphi\|_*^2 e^\varphi \, d\mu. \tag{8.9}$$

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

$$\|e^{Q_t\varphi}\|_{a+\rho' t} \leq \|e^\varphi\|_a. \tag{8.10}$$

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 8.11. *The implication (ii) $\Rightarrow$ (i) is a variant of a well-known result due to Otto and Villani [49] 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 [7]. On the other hand, in the classical setting, the equivalence (i) $\iff$ (ii) was studied and developed in [26, 27, 28].*

Observe that the relations between the various constants are almost optimal. Indeed, starting from $\overline{\mathbf{T}}_2^-(b)$, we deduce from (ii) $\Rightarrow$ (iii) that the log-Sobolev inequality (8.9) 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 make use of the above result to obtain examples of measures satisfying $\overline{\mathbf{T}}_2^-(b)$ in Section 8.3. Indeed, the "convex" log-Sobolev inequality (8.9) was studied in the literature [1].

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

Lemma 8.12. *Let $\|\cdot\|$ be a norm on $\mathbb{R}^m$ whose dual norm is strictly convex. If $\varphi : \mathbb{R}^m \to \mathbb{R}$ is a convex function, then for all $t > 0$, the function $Q_t\varphi$ defined by*

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

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

$$(Q_t\varphi)^*(x) = \sup_{y\in\mathbb{R}^m} \{x\cdot y - Q_t\varphi(y)\}, \qquad x \in \mathbb{R}^m.$$

A simple calculation shows that $(Q_t\varphi)^*(x) = \varphi^*(x) + \frac{t}{2}\|x\|_*^2$, for all $x \in \mathbb{R}^m$. By assumption, $\|\cdot\|_*$ satisfies (8.6). This easily implies that (and is in fact equivalent to) 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. [35, Theorem E.4.1.1]) then implies that $(Q_t\varphi)^{**} = Q_t\varphi$ is $\mathcal{C}^1$-smooth on $\mathbb{R}^m$. □

Proof of Theorem 8.8. That (i) implies (ii) is given in Proposition 8.3.

To prove that (ii) implies (iii), fix $\varphi\colon \mathbb{R}^m \to \mathbb{R}$ 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

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

Hence, given $\lambda > 0$ and $x \in \mathbb{R}^m$, by the Young inequality $u\cdot v \le \frac{1}{2\lambda}\|u\|_*^2 + \frac{\lambda}{2}\|v\|^2$, $u, v \in \mathbb{R}^m$, we have

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

The expected result follows.

To prove that (iii) implies (iv), we follow the now classical argument from [7] based on the Hamilton-Jacobi equation satisfied by $(t,x) \mapsto Q_t\varphi(x)$. Since we do not assume that μ 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 [7, Theorem 2.1] to our framework. First, as shown in [28] or [4], the following Hamilton-Jacobi equation holds for *all* $t > 0$ and $x \in \mathbb{R}^m$:

$$\frac{d^+}{dt}Q_t\varphi(x) = -\frac{1}{2}|\nabla^- Q_t\varphi|^2(x), \tag{8.13}$$

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

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

Here, since φ is *convex*, the regularization property of the inf-convolution operator Q_t given in Lemma 8.12 implies that for all $t > 0$, the function $x \mapsto Q_t f(x)$ 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 8.12 again, if $\varphi\colon \mathbb{R}^m \to \mathbb{R}$ is convex, then so is $Q_t\varphi$. Therefore, (8.9) 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 [7, Theorem 2.1] (see also [28, Theorem 1.11]).

Finally we prove that (iv) implies (i). We observe that, at $t = 1$ and $a = 0$, (8.10) means precisely 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 4.5 and to the fact that, as recalled above, $Q_1\varphi = R_c\varphi = \inf_{p\in\mathcal{P}_1(X)} \left\{\int \varphi(y)\, p(dy) + c(x,p)\right\}$, for any $\varphi\colon \mathbb{R}^m \to \mathbb{R}$ convex, Lipschitz and bounded from below. This completes the proof. □

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. [67]). We recall that if $c : X \times X$ is some cost function on a space X, a function $f : X \to \mathbb{R}\cup\{\pm\infty\}$ is said to be c-convex, if there exists some function $g : X \to \mathbb{R}\cup\{\pm\infty\}$ such that

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

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 (8.6). In other words, a function $f\colon \mathbb{R}^m \to \mathbb{R}\cup\{\pm\infty\}$ is $\frac{\lambda}{2}\|\cdot\|^2$-convex, if there exists $g\colon \mathbb{R}^m \to \mathbb{R}\cup\{\pm\infty\}$ such that $f = P_{\frac{1}{\lambda}} g$ (recall the definition of P_t from 8.5). In [28, Proposition 2.2], for example, 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, 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 $\mathrm{Hess} f \geq -\lambda \mathrm{Id}$ (as a matrix), where Hess denotes the Hessian (see e.g. [28, Proposition 2.3]).

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 $f : \mathbb{R}^m \to \mathbb{R}$ that are *concave, Lipschitz, bounded from above* and $\frac{\lambda}{2}\|\cdot\|^2$-convex.

Remark 8.14. *According to Lemma 8.12, 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 8.15. *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

$$\mathrm{Ent}_\mu(e^\varphi) \leq C\int (\varphi - Q_{1/\lambda}\varphi)e^\varphi\, d\mu; \tag{8.16}$$

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

$$\mathrm{Ent}_\mu(e^\varphi) \leq \frac{1}{2\rho}\int \|\nabla\varphi\|_*^2 e^\varphi\, d\mu. \tag{8.17}$$

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 8.18. *Also, Equation (8.16) is very close to (yet different from) the (τ)-log-Sobolev inequality (8.2). 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 emphasize 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 by a factor 12, at the most.*

Proof. To prove that (i) implies (ii), we follow the argument in the proof of Proposition 8.3. 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

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

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 2.11) to ensure that (since $Q_1(-\varphi) = -P_1\varphi$)

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

Applying $\overline{\mathbf{T}}_2^+(b)$ and rearranging the terms, we end up with the following inequality, noting that $\int \exp\{P_s f\}\, d\mu H(\nu_f|\mu) = \mathrm{Ent}_\mu\left(\exp\{P_s f\}\right)$:

$$\mathrm{Ent}_\mu\left(e^{P_s f}\right) \le \frac{1}{1-\lambda b} \int [P_s f - f] e^{P_s f}\, d\mu\,,$$

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[3] (for any $s \ge 0$), 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 [28].

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. [28, Proposition 2.2]). Define $m(x) = \left\{\bar{y} \in \mathbb{R}^m : f(x) = Q_s f(\bar{y}) - \frac{\lambda'}{2}\|x-\bar{y}\|^2\right\}$, *i.e.* the set of points where the supremum is reached, which is non-empty by simple compactness arguments (see [28, Lemma 2.6]). Given $\bar{y} \in m(x)$, we have for all $z \in \mathbb{R}^m$,

$$f(x) = Q_s f(\bar{y}) - \frac{\lambda'}{2}\|x-\bar{y}\|^2 \le f(z) + \frac{\lambda'}{2}\left(\|z-\bar{y}\|^2 - \|x-\bar{y}\|^2\right)\,. \tag{8.19}$$

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

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

Inserting this inequality in (8.19), one gets

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

[3]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, $Q_s f \le f$ and therefore is bounded above.

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

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

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

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

According to (8.19), the triangle inequality, and the inequality $\|x - \bar{y}\| \le \frac{1}{\lambda'}\|\nabla f(x)\|_*$, one gets

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

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

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

which inserted into (8.16) yields (8.17).

It remains to prove that (iii) implies (i). To that purpose, let $\ell(t) := -\rho(1-t)$, $t \in (0,1)$ (observe that $\ell(t) \le 0$), set $s = -\ell(t)/\lambda'$, and consider a function $f\colon \mathbb{R}^m \to \mathbb{R}$, convex, Lipschitz and bounded below. 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 φ is concave, Lipschitz, bounded above and $\lambda' c$-convex. Since f is convex, $Q_t f$ is convex and so, since $\ell(t) \le 0$, φ is concave. On the other hand, since f is Lipschitz, so is φ. Also, f being bounded below, $Q_t f \ge \inf f$ and $\ell(t) \le 0$, we have $\varphi = \ell(t)Q_t f \le \ell(t)\inf f$ which proves that φ 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 $s \le t$),

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

hence φ is $\lambda' c$-convex. In turn, applying the log-Sobolev inequality to φ (which is $\mathcal{C}^1$-smooth according to Lemma 8.12), we end up with the following inequality that we shall use later on:

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

where $H(t) := \int e^{\ell(t)Q_t f} d\mu$. Hence, by the Hamilton-Jacobi equation (8.13),

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

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

$$\|e^{Q_1 f}\|_{\ell(1)} \le \left\|e^{Q_{\frac{\rho}{\rho+\lambda'}} f}\right\|_{\ell(\frac{\rho}{\rho+\lambda'})}$$

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 4.5 and Corollary 2.11. This completes the proof. □

8.3. **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 8.20 ([1]). *Let μ 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 $\varphi\colon \mathbb{R} \to \mathbb{R}$, it holds*

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

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

Corollary 8.21. *Let μ 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 8.20 guarantees that Item (iii) of Theorem 8.8 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 8.8. □

We refer to [29] for a complete characterization of the inequalities $\overline{\mathbf{T}}_2^\pm$ (and other $\overline{\mathbf{T}}$ inequalities) on the line. As proved there in [29, Theorem 1.2], a probability measure μ satisfies $\overline{\mathbf{T}}^-(C)$ and $\overline{\mathbf{T}}^+(C)$ for some C if and only if there exists some $D > 0$ such that the monotone increasing rearrangement map U transporting the symmetric exponential probability measure $\nu(dx) = \frac{1}{2}e^{-|x|}\,dx$ on μ satisfies the following growth condition:

$$\sup_x |U(x+u) - U(x)| \leq \frac{1}{D}\sqrt{u+1}, \qquad \forall u > 0.$$

See [29] for an explicit relation between the constants C and D, and also for more general statements.

9. Generalization of Kantorovich duality

9.1. **Notations and statements of the duality results.** First let us recall and complete the notations introduced in Section 2.1. Throughout this section, (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 $\mathcal{M}(X)$.

If $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ is a continuous function satisfying (2.1), 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) $x_o \in X$.

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 we recall that $\Phi_\gamma(X)$ denotes the set of continuous functions $\varphi : X \to \mathbb{R}$ satisfying the growth condition (2.2). 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

$$U_{\varphi,a,\varepsilon} := \left\{ m \in \mathcal{M}_\gamma(X); \left| \int \varphi \, dm - a \right| < \varepsilon \right\}, \ \varphi \in \Phi_\gamma(X), \ a \in \mathbb{R}, \ \varepsilon > 0. \tag{9.1}$$

The set $\mathcal{P}_\gamma(X) := \mathcal{P}(X) \cap \mathcal{M}_\gamma(X)$ is equipped with the trace topology denoted by $\sigma(\mathcal{P}_\gamma(X))$. Let us remark that if γ is bounded, then $\mathcal{P}_\gamma(X) = \mathcal{P}(X)$ and the topology $\sigma(\mathcal{P}_\gamma(X))$ is the usual weak topology on $\mathcal{P}(X)$.

The following lemma clarifies the link between the topology defined by the class $\Phi_\gamma(X)$ and the usual weak convergence topology (defined by bounded continuous functions).

Lemma 9.2. *Let $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function satisfying (2.1) and define $\xi(x) = 1 + \gamma(d(x_o, x))$, $x \in X$, where x_o is some point in X.*

- *(i) Let $(m_n)_{n\geq 1}$ be a sequence in $\mathcal{M}_\gamma(X)$. The following are equivalent*
  - *(a) m_n converges to $m \in \mathcal{M}_\gamma(X)$ for the topology $\sigma(\mathcal{M}_\gamma(X))$,*
  - *(b) ξm_n converges to ξm for the usual weak convergence topology in $\mathcal{M}(X)$.*
- *(ii) Let $(\mu_n)_{n\geq 1}$ be a sequence in $\in \mathcal{P}_\gamma(X)$. The following are equivalent*
  - *(a) μ_n converges to $\mu \in \mathcal{P}_\gamma(X)$ for the topology $\sigma(\mathcal{P}_\gamma(X))$,*
  - *(b) μ_n converges to μ for the usual weak convergence topology in $\mathcal{P}(X)$ and*
  $$\int \gamma(d(x_o, x)) \, \mu_n(dx) \to \int \gamma(d(x_o, x)) \, \mu(dx)$$
  for some (and hence all) $x_o \in X$.
- *(iii) The topology $\sigma(\mathcal{P}_\gamma(X))$ on $\mathcal{P}_\gamma(X)$ is separable and coincides with the topology defined by the metric d_γ defined by*

$$d_\gamma(\mu, \nu) = \sup \left\{ \int f \, \xi d\mu - \int f \xi \, d\nu; f \ 1-\text{Lipschitz and } \sup_{x\in X} |f(x)| \leq 1 \right\}, \qquad \mu, \nu \in \mathcal{P}_\gamma(X).$$

Of course, the preceding result is easily adapted to $\mathcal{M}_\gamma(X \times X)$ equipped with $\sigma(\mathcal{M}_\gamma(X \times X))$ (resp. $\mathcal{P}_\gamma(X \times X)$ equipped with $\sigma(\mathcal{P}_\gamma(X \times X))$). The proof of Lemma 9.2 is postponed to the end of this section.

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 $\varphi : X \times X \to \mathbb{R}$ 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 $x, y \in X$.

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

Before stating our main result, we need to introduce some technical assumptions and comment on them. Below we denote by $\pi(dxdy) = p_x(dy)\pi_1(dx)$ the disintegration of a probability measure π on $X \times X$ with respect to its first marginal π_1.

Definition 9.3 (Conditions (C), (C'), (C''))**.** *Given (X, d) a complete separable metric space, $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ a continuous function satisfying (2.1) and $c\colon X \times \mathcal{P}_\gamma(X) \to [0, \infty]$, a cost function, we say the condition (C) holds if*

- *(C_1) For all $\mu \in \mathcal{P}_\gamma(X)$, the function $\pi \mapsto I_c[\pi] := \int c(x, p_x) \, \pi_1(dx)$ 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_2) The function $p \mapsto c(x, p)$ is convex for all $x \in X$.*
- *(C_3) The function $(x, p) \mapsto c(x, p)$ is continuous with respect to the product topology.*
- *(C_4) 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_3') *(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_3'') *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 9.6, 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 9.6. In particular, such a result applies to the transport costs $\widetilde{\mathcal{T}}$, $\overline{\mathcal{T}}$ and $\widehat{\mathcal{T}}$ introduced in Section 2.4.

Proposition 9.4. *Let (X,d) be a complete separable metric space. Let $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function satisfying (2.1) and let $(\varphi_k)_{k\in\mathbb{N}}$ be a sequence of elements of $\Phi_\gamma(X \times X)$ such that $\varphi_0 \equiv 0$. Assume that the cost function $c : X \times \mathcal{P}_\gamma(X) \to [0,\infty]$ is defined by*

$$c(x,p) = \sup_{k\in\mathbb{N}} \int \varphi_k(x,y)\, p(dy), \qquad \forall x \in X, \qquad \forall p \in \mathcal{P}_\gamma(X). \tag{9.5}$$

Then Conditions (C_1) and (C_2) hold and $c : X \times \mathcal{P}_\gamma(X) \to [0,\infty]$ 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 9.6. *Let (X,d) be a complete separable metric space. Let $\gamma\colon \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function satisfying (2.1) and $c\colon X \times \mathcal{P}_\gamma(X) \to [0,\infty]$ be a cost function. 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 9.7. *Note that since $c \geq 0$, $R_c\varphi$ is bounded from below as soon as φ 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 since X is compact in this case.*

The proof of Theorem 9.6 uses classical tools from convex analysis that we recall in a separate subsection (see Section 9.2 below), and then apply them to our specific setting. We refer to Mikami [45], Léonard [38], Tan-Touzi [65] for similar strategies.

Proof of Lemma 9.2. Since γ is continuous, the function ξ is also continuous and so $\Phi_\gamma(X) = \{f\xi; f \in \mathcal{C}_b(X)\}$. Therefore a sequence $(m_n)_n$ converges to m in $\mathcal{M}_\gamma(X)$ if and only if $\xi m_n \to \xi m$ for the usual weak convergence, which proves *(i)*. The proof of Item *(ii)* can be easily adapted from the one of *e.g* [66, Theorem 7.12]. Let us turn to the proof of Item *(iii)*. According to *e.g* [13, Theorem 8.3.2], the usual weak convergence topology on $\mathcal{M}^+(X)$ (the set of non-negative finite

Borel measures on X) coincides with the one generated by the so called Fortet-Mourier distance d defined by

$$d(m_1, m_2) = \sup\left\{\int f\,dm_1 - \int f\,dm_2; f\ 1-\text{Lipschitz and } \sup_{x\in X}|f(x)| \leq 1\right\}, \qquad m_1, m_2 \in \mathcal{M}^+(X).$$

First observe that $d_\gamma(\mu,\nu) = d(\xi\mu,\xi\nu)$, $\mu,\nu \in \mathcal{P}_\gamma(X)$, from which easily follows that d_γ is a distance on $\mathcal{P}_\gamma(X)$. According to point *(i)*, one concludes that

$$\mu_n \to \mu \text{ in } \mathcal{P}_\gamma(X) \ \Leftrightarrow\ \xi\mu_n \to \xi\mu \text{ in } \mathcal{M}^+(X) \ \Leftrightarrow\ d(\xi\mu_n,\xi\mu) = d_\gamma(\mu_n,\mu) \to 0,$$

which completes the proof of the metrizability. Finally, let us prove the separability of the topology $\sigma(\mathcal{P}_\gamma(X))$. Let $E \subset X$ be a countable dense subset and let us show that the set

$$F = \left\{\frac{1}{n}\sum_{i=1}^n \delta_{x_i} : n \geq 1, x_1,\ldots,x_n \in E\right\}$$

is a (countable) dense subset of $\mathcal{P}_\gamma(X)$. Let $(X_i)_{i\geq1}$ be an i.i.d sequence of law μ and define $L_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. According to Varadarajan's theorem, with probability 1, the sequence L_n converges to μ for the usual weak topology of $\mathcal{P}(X)$. Moreover, according to the strong law of large numbers, for any $x_o \in X$, $\int \gamma(d(x_o,x))\,L_n(dx) \to \int \gamma(d(x_o,x))\,\mu(dx)$ with probability 1. According to Item *(ii)*, this shows in particular that there exists at least one sequence $(x_i)_{i\geq1}$ such that $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ converges to μ for the topology $\sigma(\mathcal{P}_\gamma(X))$. Now, observe that the map $z = (z_1,\ldots,z_n) \mapsto L_n^z := \frac{1}{n}\sum_{i=1}^n \delta_{z_i}$ is continuous. Namely, for all $\varphi \in \Phi_\gamma(X)$, the function $z \mapsto \int \varphi(y)\,L_n^z(dy) = \frac{1}{n}\sum_{i=1}^n \varphi(z_i)$ is clearly continuous on X^n. Therefore, since E is dense in X, for all $n \geq 1$ there exists $y^{(n)} \in E^n$ such that $d_\gamma(\mu_n, L_n^{y^{(n)}}) \leq 1/n$. From this follows that $\tilde{\mu}_n := L_n^{y^{(n)}}$ is a sequence of elements of F converging to μ which completes the proof. □

9.2. **Fenchel-Legendre duality.** The main tool used in the proof of Theorem 9.6 is the following Fenchel-Legendre duality theorem (see for instance [68, Theorem 2.3.3]).

Theorem 9.8 (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 $F\colon E \to]-\infty,\infty]$, it holds*

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

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

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

To apply Theorem 9.8 in our framework, one needs to identify the topological dual space of $(\mathcal{M}_\gamma(X), \sigma(\mathcal{M}_\gamma(X)))$ equipped with the topology defined in Section 9.1. 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 9.9. *A linear form $\ell\colon \mathcal{M}_\gamma(X) \to \mathbb{R}$ 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*

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

The proof of this lemma appears, for instance, in the book by Deuschel and Stroock [18]. We recall it here for the sake of completeness.

Proof of Lemma 9.9. 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 ℓ be a continuous linear functional and let us show that ℓ is of the preceding form. Define $\varphi(x) = \ell(\delta_x)$, $x \in X$ (where δ_x is the Dirac mass at x). First we will show that φ belongs to $\Phi_\gamma(X)$. The map $X \ni x \mapsto \delta_x \in \mathcal{M}_\gamma(X)$ is continuous. Namely, for all $\varphi_1,\ldots,\varphi_n \in \Phi_\gamma$, 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 (9.1)) and this set is open, which proves that $x \mapsto \delta_x$ is continuous on X. As a result φ is continuous. It remains to prove that φ satisfies the growth condition (2.2). Since ℓ is continuous, the set $O := \{m \in \mathcal{M}_\gamma(X); |\ell(m)| < 1\}$ is open and contains 0. By the definition of the topology $\sigma(\mathcal{M}_\gamma(X))$, there exist an integer n, $\varphi_1, \ldots, \varphi_n \in \Phi_\gamma$, $a_1, \ldots, a_n \in \mathbb{R}$ and $\varepsilon_1, \ldots, \varepsilon_n > 0$ 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,

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

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

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

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

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

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

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 (see *e.g* the argument given in [13, Example 8.1.6]), the equality $\ell(m) = \int \varphi \, dm$ extends to any $m \in \mathcal{M}_\gamma(X)$. □

9.3. **Compactness and primal attainment.** During the proof of Theorem 9.6, we will also use the following easy extension of Prokhorov's theorem.

Theorem 9.10. *Let $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function satisfying (2.1) and define $\xi(x) = 1 + \gamma(d(x_o, x))$, $x \in X$, where x_o is some arbitrary fixed point in X. A set $A \subset \mathcal{P}_\gamma(X)$ is relatively compact for the topology $\sigma(\mathcal{P}_\gamma(X))$ if and only if for all $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subset X$ such that*

$$\sup_{\nu \in A} \int_{X \setminus K_\varepsilon} \xi(x) \, \nu(dx) \leq \varepsilon. \tag{9.11}$$

A similar result holds of course on $\mathcal{P}_\gamma(X \times X)$ equipped with the topology $\sigma(\mathcal{P}_\gamma(X \times X))$.

Proof. It follows easily from Item *(i)* of Lemma 9.2 that a set $A \subset \mathcal{P}_\gamma(X)$ is relatively compact for the topology $\sigma(\mathcal{P}_\gamma(X))$ if and only if $\tilde{A} = \{\xi\nu \,;\, \nu \in A\}$ is relatively compact in $\mathcal{M}^+(X)$ (the set of finite non-negative Borel measures on X) for the usual weak convergence topology. According to the classical version of Prokhorov theorem (see *e.g* [13, Theorem 8.6.2]) this latter condition holds if and only if *(i)* $\sup_{\nu \in A} \int \xi(x)\nu(dx) < +\infty$ and *(ii)* for all $\varepsilon > 0$ there exists a compact $K_\varepsilon \subset X$ such that (9.11) holds. Since ξ is continuous, it is bounded on K_1 and so *(i)* is a consequence of *(ii)* (with $\varepsilon = 1$), which completes the proof. □

Before presenting the proof of Theorem 9.6, let us mention a simple consequence of Theorem 9.10 in terms of existence of optimal coupling for our generalized transport costs.

Corollary 9.12. *Let (X, d) be a complete separable metric space. Let $\gamma\colon \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function and $c\colon X \times \mathcal{P}_\gamma(X) \to [0, \infty]$ be a cost function. Assume that the cost function c satisfies Condition (C_1). Then for all $\mu, \nu \in \mathcal{P}_\gamma(X)$ such that $\mathcal{T}_c(\nu|\mu) < \infty$ there exists a coupling π^*, such that $\mathcal{T}_c(\nu|\mu) = I_c(\pi^*)$.*

Proof. The proof is easily adapted from [67, Theorem 4.1]. Details are left to the reader. □

9.4. **Proof of Theorem 9.6 (Duality).**

Proof of Theorem 9.6. Fix $\mu \in \mathcal{P}_\gamma(X)$ and let us consider the function F defined on $\mathcal{M}_\gamma(X)$ by

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

Let us show that the function F satisfies the assumptions of Theorem 9.8.

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 μ almost every $x \in X$. From condition (C_2), the cost function c is convex in its second argument, and therefore it holds

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

Optimizing over π_0, π_1 gives $F((1-t)\nu_0 + t\nu_1) \le (1-t)F(\nu_0) + tF(\nu_1)$, 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 $F(m) \le \liminf_{n\to\infty} F(m_n)$. One can assume without loss of generality that $F(m_n) < \infty$ for all n. By the definition of $\mathcal{T}_c(\,\cdot\,|\mu)$, for all $n \in \mathbb{N}^*$, there exists $\pi_n \in \Pi(\mu, m_n)$ such that $I_c[\pi_n] - 1/n \le \mathcal{T}_c(m_n|\mu) \le I_c[\pi_n]$. Since m_n is a converging sequence, the set $\{m_n; n \in \mathbb{N}^*\} \cup \{\mu\}$ is relatively compact. Therefore, according to Theorem 9.10, for some arbitrary fixed point $x_o \in X$, for all $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subset X$ such that

$$\sup_{n\in\mathbb{N}^*} \int_{X\setminus K_\varepsilon} 1 + \gamma(d(x_o, y))\, m_n(dy) \le \varepsilon$$

and

$$\int_{X\setminus K_\varepsilon} 1 + \gamma(d(x_o, x))\mu(dx) \le \varepsilon.$$

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

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

So according to Theorem 9.10, 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 π_n converges to some $\pi^* \in \mathcal{P}_\gamma(X \times X)$. This π^* has the correct marginals μ and m. Furthermore, denoting by $\ell = \liminf_{n\to\infty} I_c[\pi_n] = \liminf_{n\to\infty} \mathcal{T}_c(m_n|\mu)$, we see that, for all $r > 0$,

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

for all but finitely 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 π^* also belongs to $A_{\ell+r}$. In other words,

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

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

According to Lemma 9.9, 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 9.8 together with Lemma

9.9 we conclude that, for any $m \in \mathcal{P}_\gamma(X)$,

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

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

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

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

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

Therefore,

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

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

$$F^*(-\varphi) = -\int R_c\varphi(x) \, \mu_*(dx), \qquad \forall \varphi \in \Phi_{\gamma,b}(X). \tag{9.13}$$

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

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

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

$$F^*(-\varphi) \le -\int R_c\varphi(x) \, \mu_*(dx).$$

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^ε defined by

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

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

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

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

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

When the condition (C_3) holds, the kernel p_x^ε is obtained by applying the elementary measurable selection result of Lemma 9.14 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 and separable according to Item *(iii)* of Lemma 9.2.

Under condition (C_3'), the space $Y = \mathcal{P}_\gamma(X)$ is compact and the function H defined above is lower-semicontinuous. The selection Lemma 9.15 below ensures that there exists a *measurable* kernel $X \to \mathcal{P}_\gamma(X) : x \mapsto p_x$ 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^ε in the non-empty set M_x^ε, we get a measurable kernel $X \to \mathcal{P}_\gamma(X) : x \mapsto p_x^\varepsilon$. 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 π for the topology $\sigma(\mathcal{P}_\gamma(X\times X))$. Write $\pi_n(dxdy) = p_{x,n}(dy)\,\mu(dx)$ and denote by ν_n (resp. ν) the second marginal of π_n (resp. π). The sequence $(\nu_n)_n$ converges to ν, therefore it is relatively compact and so according to Theorem 9.10, for all $\varepsilon > 0$, there is some compact $K_\varepsilon \subset X$ (i.e. a finite set) such that $\int_{K_\varepsilon^c} 1 + \gamma(d(x_o,y))\,\nu_n(dy) \le \varepsilon$, where x_o is some fixed point in X. In other words,

$$\sum_{y\in K_\varepsilon^c}\sum_{x\in X}(1+\gamma(d(x_o,y)))\,p_{x,n}(\{y\})\mu(\{x\}) \le \varepsilon\,.$$

In particular, for all $x \in X$ in the support of μ, it holds

$$\sum_{y\in K_\varepsilon^c}(1+\gamma(d(x_o,y)))\,p_{x,n}(\{y\}) \le \varepsilon/\mu(\{x\}),$$

and so, according to Theorem 9.10, $\{p_{x,n}; n\in\mathbb{N}\}$ 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 μ, $\{p_{x,n}; n\in\mathbb{N}\}$ is relatively compact, the classical diagonal extraction argument enables us to construct an increasing map $\sigma:\mathbb{N}\to\mathbb{N}$ such that $\tilde p_{x,\sigma(n)}$ converges to some $p_x \in \mathcal{P}_\gamma(X)$ as $n\to\infty$, for all x in the support of μ. Finally, using Fatou's lemma and the lower-semicontinuity of $p\mapsto c(x,p)$, one gets

$$\begin{aligned}\lim_{n\to\infty} I_c[\pi_n] &= \lim_{n\to\infty}\int c(x,p_{x,\sigma(n)})\,\mu(dx)\\ &\ge \int \liminf_{n\to\infty} c(x,p_{x,\sigma(n)})\,\mu(dx)\\ &\ge \int c(x,p_x)\,\mu(dx).\end{aligned}$$

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

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

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

In the proof of Theorem 9.6 we used the following results, elementary proofs of which can be found in [5] (see Proposition 7.34 and Proposition 7.33).

Lemma 9.14. *Let X be a metrizable space, Y a metrizable and separable space and $H\colon X\times Y\to \mathbb{R}\cup\{+\infty\}$ 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*

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

Lemma 9.15. *Let X be a metrizable space, Y a compact metrizable space and $H\colon X \times Y \to \mathbb{R} \cup \{+\infty\}$ be a lower-semicontinuous function. Then there exists a measurable function $x \mapsto s(x)$ such that for all $x \in X$,*

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

9.5. Proofs of Theorems 2.7, 2.11 and 2.14.

9.5.1. *Proof of the usual Kantorovich duality theorem.* As a warm up, let us begin with the proof of the classical Kantorovich duality that we restate below. Recall that $\Phi_{\gamma_0}(X)$ (that corresponds to Φ_γ, defined in Section 2.1, with the special choice $\gamma = \gamma_0$) stands for the set of continuous and bounded function that is usually denoted by $\mathcal{C}_b(X)$.

Theorem 9.16. *Let (X, d) be a complete separable metric space. Assume that $\omega : X \times X \to [0, \infty]$ is some lower-semicontinuous cost function. Then it holds,*

$$\mathcal{T}_\omega(\nu, \mu) = \sup_{\varphi \in \mathcal{C}_b(X)} \left\{ \int Q_\omega \varphi(x)\, \mu_*(dx) - \int \varphi(y)\, \nu(dy) \right\}, \quad \mu, \nu \in \mathcal{P}(X), \tag{9.17}$$

where μ_* denotes the inner measure induced by μ and

$$Q_\omega \varphi(x) = \inf_{y \in X} \{\varphi(y) + \omega(x, y)\}, \qquad x \in X, \quad \varphi \in \mathcal{C}_b(X).$$

Proof of Theorem 9.16. First assume that $\omega\colon X \times X \to [0, \infty)$ 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 $I_c[\pi] = \int \omega(x, y)\, \pi(dxdy)$ (recall the notation I_c from Section 2.2) 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 9.6 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 [66, Item 3 in the proof of Theorem 1.3]. □

9.5.2. *Proof of Theorem 2.7.*

Proof of Theorem 2.7. Depending on the assumption on the space and on α, one needs to verify that Condition (C), (C') or (C'') of Theorem 9.6 is satisfied. We distinguish between the different cases.

Assume first that $\alpha\colon \mathbb{R}_+ \to \mathbb{R}_+$ is convex and continuous and $\delta\colon \mathbb{R}_+ \to \mathbb{R}_+$ is continuous. Let $\gamma = \delta$. Then the cost $c(x, p) = \alpha\left(\int \delta(d(x, y))\, p(dy)\right)$ is clearly convex with respect to p. Let us show that it is continuous on $X \times \mathcal{P}_\gamma(X)$ (equipped with the product topology). Suppose that (x_n, p_n) is some sequence converging to (x, p), then it holds

$$\begin{aligned} &\left| \int \delta(d(x_n, y))\, p_n(dy) - \int \delta(d(x, y))\, p(dy) \right| \\ &\le \int |\delta(d(x_n, y)) - \delta(d(x, y))|\, p_n(dy) + \left| \int \delta(d(x, y))\, p_n(dy) - \int \delta(d(x, y))\, p(dy) \right|. \end{aligned}$$

By definition of the topology $\sigma(\mathcal{P}_\gamma(X))$, the second term in the right hand side goes to 0 as n goes to infinity. Let us show that the first term also tends to 0. Since p_n is a converging sequence, according to Theorem 9.10, for any $\varepsilon > 0$ there is a compact set K_ε such that $\int_{X\setminus K_\varepsilon} 1+\delta(d(x,y))\, p_n(dy) \leq \varepsilon$. So

$$\int |\delta(d(x_n,y)) - \delta(d(x,y))|\, p_n(dy) \leq \sup_{y\in K_\varepsilon} |\delta(d(x_n,y)) - \delta(d(x,y))| \\ + C\delta(d(x_n,x))p_n(X\setminus K_\varepsilon) + (C+1)\int_{X\setminus K_\varepsilon} \delta(d(x,y))\, p_n(dy),$$

where C is the constant appearing in (2.1). Therefore, thanks to the compactness of K_ε, it is easily seen that

$$\limsup_{n\to\infty} \int |\delta(d(x_n,y)) - \delta(d(x,y))|\, p_n(dy) \leq (C+1+C\delta(0))\varepsilon$$

and letting $\varepsilon \to 0$ completes the proof.

Assumptions $(C_2), (C_3)$ of Theorem 9.6 are thus fulfilled. Let us check that Condition (C_4) holds. By Jensen inequality, if $\int \alpha\left(\int \delta(d(x,y))\, p_x(dy)\right)\mu(dx) < +\infty$ for some $\mu \in \mathcal{P}_\gamma(X)$, then $\iint \delta(d(x,y))\, p_x(dy)\mu(dx) < \infty$. Therefore, if $x_o \in X$ is some fixed point, using (2.1) yields to

$$\begin{aligned} \int d(x_o,y)(\mu p)(dy) &= \iint \delta(d(x_o,y))\, p_x(dy)\mu(dx) \\ &\leq C\int \delta(d(x_o,x))\,\mu(dx) + C\iint \delta(d(x,y))\, p_x(dy)\mu(dx) \\ &< \infty \end{aligned}$$

and so $\mu p \in \mathcal{P}_\gamma(X)$.

As for Condition (C_1), let us set $\alpha(t) = +\infty$ for $t < 0$, so that α is lower-semicontinuous on $\mathbb{R}$. According to the Fenchel-Legendre duality (Theorem 9.8),

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

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

$$c(x,p) = \sup_{s\geq 0}\int s\delta(d(x,y)) - \alpha^*(s)\, p(dy) = \sup_{(s,t)\in \mathrm{epi}(\alpha^*)} \int s\delta(d(x,y)) - t\, p(dy) = \sup_{k\in\mathbb{N}} \int \varphi_k(x,y)\, p(dy),$$

with $\varphi_0 = 0$ and $\varphi_k(x,y) = s_k\delta(d(x,y)) - t_k$, $k \geq 1$ where $(s_k,t_k)_{k\geq 1}$ is any dense subset of $\mathrm{epi}(\alpha^*) = \{(s,t) \in [0,\infty)\times\mathbb{R}; t \geq \alpha^*(s)\}$. For all $k \in \mathbb{N}$, $\varphi_k \in \Phi_\gamma(X\times X)$ and so according to Proposition 9.4, the cost function c verifies (C_1).

Now assume that $\alpha : \mathbb{R} \to [0,+\infty]$ is convex and lower-semicontinuous and δ is lower-semicontinuous. Let $\gamma = \gamma_0$ and therefore $\mathcal{P}_{\gamma_0}(X) = \mathcal{P}(X)$, $\Phi_{\gamma_0}(X) = \mathcal{C}_b(X)$. The function c is also clearly convex with respect to p (hence Condition (C_2) is satisfied). Condition (C_4) is obvious since $\mathcal{P}_{\gamma_0}(X) = \mathcal{P}(X)$. Since δ is lower-semicontinuous, there exists an increasing sequence $(\delta_N)_{N\in\mathbb{N}}$ of Lipschitz continuous functions $\delta_N : \mathbb{R}_+ \to \mathbb{R}_+$ that converges to δ (for example $\delta_N(u) = \inf_{v\in\mathbb{R}}\{\delta(v) + N|u-v|\}$). By using the Fenchel-Legendre duality for α as above and by monotone convergence, one has

$$\begin{aligned} c(x,p) &= \sup_{(s,t)\in\mathrm{epi}(\alpha^*)} \sup_{N\in\mathbb{N}} \int s\delta_N(d(x,y)) - t\, p(dy) \\ &= \sup_{(s,t)\in\mathrm{epi}(\alpha^*)} \sup_{N\in\mathbb{N}} \sup_{M\in\mathbb{N}} \int s\left(\delta_N(d(x,y)) \wedge M\right) - t\, p(dy) \\ &= \sup_{k\in\mathbb{N}} \int \varphi_k(x,y)\, p(dy), \end{aligned}$$

with $\varphi_0 = 0$ and $\varphi_k(x,y) = s_{\ell(k)}\left(\delta_{N(k)}(d(x,y)) \wedge M(k)\right) - t_{\ell(k)}$, $k \geq 1$, where $(s_l,t_l)_{l\in\mathbb{N}}$ is any dense subset of $\mathrm{epi}(\alpha^*) = \{(s,t) \in [0,\infty)\times\mathbb{R}; t \geq \alpha^*(s)\}$, and the map $\mathbb{N}^* \ni k \mapsto (N(k), \ell(k), M(k)) \in \mathbb{N}\times\mathbb{N}\times\mathbb{N}$ is one-to-one. Since $\varphi_k \in \mathcal{C}_b(X)$ for any $k \in \mathbb{N}^*$, by Proposition 9.4, 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. The result of Theorem 2.7 is finally a direct consequence of Theorem 9.6. □

9.5.3. *Proof of Theorem 2.11.*

Proof of Theorem 2.11.
(1) The proof of the first item is similar to that of Theorem 2.7. Namely, if $\theta\colon \mathbb{R}^m \to \mathbb{R}$ 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 θ, one sees that

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

with $\mathrm{epi}(\theta^*) = \{(s,t) \in \mathbb{R}^m \times \mathbb{R}; \theta^*(s) \leq t\}$. Taking a dense countable subset $(s_k,t_k)_{k\geq 1}$ of $\mathrm{epi}(\theta^*)$, 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 9.4, the cost function c satisfies (C_1).

If $\theta\colon \mathbb{R}^m \to (-\infty, +\infty]$ 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 $x \in \mathbb{R}^m$,

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

where

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

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

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

The third inequality is a consequence of Item (2), since $\psi = \overline{\psi}$ 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 Item (3) is complete. □

9.5.4. *Proof of Theorem 2.14.* Recall that X is a metric space being either compact or a countable set of isolated points. We start with an alternative representation of $c(x,p)$ that will be useful subsequently. We recall that $c : X \times \mathcal{P}(X) \to \mathbb{R}_+$ is defined by

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

if $p \ll \mu_0$ on $X \setminus \{x\}$ and $+\infty$ otherwise, where μ_0 is a reference probability measure, $\delta : \mathbb{R}_+ \to \mathbb{R}_+$ is a lower-semicontinuous function such that $\delta(u) = 0$ if $u = 0$ and $\beta : \mathbb{R}_+ \to [0,\infty]$ is a lower-semicontinuous convex function such that $\beta(0) = 0$ and $\beta(x)/x \to \infty$ as $x \to \infty$.

Lemma 9.18. *The cost function c defined above satisfies the following duality identity:*

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

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

Proof. From an easily adapted version of Theorem B.2 in [39], it holds for any finite measure ν,

$$\sup_{h\in\mathcal{C}_b(X),h\geq 0}\left\{\int h(y)\nu(dy) - \int \beta^*(h)(y)\,\mu_0(dy)\right\} = \begin{cases}\int \beta\left(\frac{d\nu}{d\mu_0}\right)d\mu_0, & \text{if } \nu \ll \mu_0\\ +\infty, & \text{otherwise.}\end{cases}$$

Given $x \in X$, we apply this result to the measure $\nu_x(dy) = \delta(d(x,y))\,p(dy)$. Since $\delta(u) = 0$ if and only if $u = 0$, the measure ν_x is absolutely continuous with respect to μ_0 if and only if p is absolutely continuous with respect to μ_0 on $X \setminus \{x\}$. □

Proof of Theorem 2.14. First, we observe that Condition (C_2) is a simple consequence of the convexity of β and Condition (C_4) is obvious since $\mathcal{P}_\gamma(X) = \mathcal{P}(X)$ $(\gamma = \gamma_0)$. According to Lemma 9.18, it holds

$$\begin{aligned}
c(x,p) &= \sup_{h\in\mathcal{C}_b(X),h\geq 0}\left\{\int h(y)\delta(d(x,y))\,p(dy) - \int \beta^*(h)\,d\mu_0\right\}\\
&= \sup_{h\in\mathcal{C}_b(X),h\geq 0}\ \sup_{N\in\mathbb{N}}\left\{\int h(y)\delta_N(d(x,y))\,p(dy) - B^*(h)\right\},\\
&= \sup_{h\in\mathcal{C}_b(X),h\geq 0}\ \sup_{N\in\mathbb{N}}\ \sup_{M\in\mathbb{N}}\left\{\int h(y)\left(\delta_N(d(x,y))\wedge M\right)\,p(dy) - B^*(h)\right\},
\end{aligned}$$

where $(\delta_N)_{N\in\mathbb{N}}$ is (as in the proof of Corollary 2.7) an increasing sequence of Lipschitz continuous functions converging to δ and $B^*(h) = \int \beta^*(h)d\mu_0$.

For all $h \in \mathcal{C}_b(X)$ non-negative, and all $N, M \in \mathbb{N}$, the function $(x,y) \mapsto h(y)\left(\delta_N(d(x,y))\wedge M\right)$ is continuous and bounded. Therefore, $p \mapsto \int h(y)\left(\delta_N(d(x,y))\wedge M\right)\,p(dy)$ is a continuous function on $X\times\mathcal{P}(X)$. Being a supremum of continuous functions, c is lower-semicontinuous on $X\times\mathcal{P}(X)$. 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 $\mathcal{C}_b(X)$ of continuous bounded functions (equipped with the norm $\|\cdot\|_\infty$) on X is separable (see [5, Proposition 7.7]). Let $\{h_\ell, \ell\in\mathbb{N}\}$ be a countable dense subset of $\mathcal{C}_b(X)$. Since $\lim_{x\to\infty}\beta(x)/x = +\infty$, β^* is convex and finite on $\mathbb{R}$, it is also continuous on $\mathbb{R}$. Therefore, the function $\mathcal{C}_b(X) \to \mathbb{R} : h \mapsto B^*(h)$ is continuous. It follows that

$$c(x,p) = \sup_{k\in\mathbb{N}}\int \varphi_k(x,y)\,p(dy), \qquad \forall x\in X, \qquad p\in\mathcal{P}(X)\,, \tag{9.19}$$

where $\varphi_0 = 0$ and $\varphi_k(x,y) = h_{\ell(k)}(y)\left(\delta_{N(k)}(d(x,y)) \wedge M(k)\right) - B^*(h_\ell(k))$, $k \geq 1$, and $\mathbb{N}^* \ni k \mapsto (\ell(k), N(k), M(k)) \in \mathbb{N} \times \mathbb{N} \times \mathbb{N}$ is one-to-one. Since, for all $k \in \mathbb{N}$, the function φ_k belongs to $\mathcal{C}_b(X,X)$, the lower-semicontinuity of I_c follows from Proposition 9.4.

Theorem 2.14 now follows from Theorem 9.6. □

9.6. **Proof of Proposition 9.4.** The proof of Proposition 9.4 is adapted from [3, Theorem 2.34].

Proof of Proposition 9.4. The function $p \mapsto c(x,p)$ is convex as a supremum of convex 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 ∞, $c_n(x,p)$ is a nondecreasing sequence converging to c. Let $\pi \in \Pi(\mu,\cdot)$, $\pi(dxdy) = p_x(dy)\mu(dx)$ such that (9.5) holds for μ-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 π. We will now prove such a property.

For μ-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 $f_0 + \cdots + f_n \leq 1$. Let us admit this claim for a moment and finish the proof of the proposition. For all $f_0,\ldots,f_n$ 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 π .

It remains to prove the claim. Obviously, if $f_0, f_1, \ldots, f_n$ 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 for all $k \leq n$, $A_k := \{x \in X; [\psi_k]_+(x) = \sup_j[\psi_j]_+(x)\}$, and define recursively $B_0 = A_0$, $B_k = A_k \setminus (B_0 \cup \cdots \cup B_{k-1})$. 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 ψ_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 [50, 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 $f_{k,\delta} : X \to [0,1]$ defined by

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

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 $\delta \to 0$, we see that

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

So if δ 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 ε 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. □

Appendix A. Proof of Theorem 4.11

The proof of the tensorization property for transport-entropy inequalities 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 A.1 (Chain rule inequality for the transport cost). *Let $\gamma\colon \mathbb{R}_+ \to \mathbb{R}_+$ be a continuous function satisfying (2.1), (X_1, d_1), (X_2, d_2) be complete separable metric spaces equipped with cost functions $c_i : X_i \times \mathcal{P}_\gamma(X_i) \to [0, \infty]$, $i \in \{1, 2\}$ 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 $c\colon X_1 \times X_2 \times \mathcal{P}_\gamma(X_1 \times X_2) \to [0, \infty)$ 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^ε 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 ν_1 and ν'_1 are the first marginals of ν, ν' respectively; the kernels $x_1 \mapsto \nu_2(x_1,\,\cdot\,)$ and $y_1 \mapsto {\nu'}_2(y_1,\,\cdot\,)$ are such that

$$\nu(dx_1dx_2) = \nu_1(dx_1)\nu_2(x_1, dx_2) \text{ and } \nu'(dy_1dy_2) = \nu'_1(dy_1)\nu'_2(y_1, dy_2);$$

and the kernel p_1^ε, defined so that $\pi_1^\varepsilon(dx_1dy_1) := \nu_1(dx_1)p_1^\varepsilon(x_1, dy_1) \in \Pi(\nu_1, \nu'_1)$, 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 A.2. *If one assumes that the cost functions c_1 and c_2 satisfy assumption (C_1) (see Definition 9.3), then the error term ε 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 9.10 above). 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 A.1. 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 ν and ν'.

To that purpose, let p_1 be a probability kernel (that depends on ε although not explicitly stated for simplicity) so that $\pi_1(dx_1dy_1) := \nu_1(dx_1)p_1(x_1, dy_1) \in \Pi(\nu_1, \nu_1')$ and

$$\int_{X_1 \times X_1} c_1(x_1, p_1(x_1,\,\cdot\,))\,\nu_1(dx_1) \le \mathcal{T}_{c_1}(\nu_1'|\nu_1) + \varepsilon. \tag{A.3}$$

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 ε) satisfying $\pi_2^{x_1,y_1}(dx_2dy_2) := \nu_2(x_1, dx_2)q_2^{x_1,y_1}(x_2, dy_2) \in \Pi(\nu_2(x_1,\,\cdot\,), \nu_2'(y_1,\,\cdot\,))$ and

$$\int_{X_2 \times X_2} c_2(x_2, q_2^{x_1,y_1}(x_2,\,\cdot\,))\,\nu_2(x_1, dx_2) \le \mathcal{T}_{c_2}(\nu_2'(y_1,\,\cdot\,)|\nu_2(x_1,\,\cdot\,)) + \varepsilon\,. \tag{A.4}$$

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

$$\begin{aligned}
&\int f(y_1, y_2)p_1(x_1, dy_1)q_2^{x_1,y_1}(x_2, dy_2)\,\nu(dx_1dx_2)\\
&\qquad = \int f(y_1, y_2)p_1(x_1, dy_1)q_2^{x_1,y_1}(x_2, dy_2)\,\nu_2(x_1, dx_2)\nu_1(dx_1)\\
&\qquad = \int f(y_1, y_2)p_1(x_1, dy_1)\nu_2'(y_1, dy_2)\nu_1(dx_1)\\
&\qquad = \int f(y_1, y_2)\nu_2'(y_1, dy_2)\nu_1'(dy_1)\\
&\qquad = \int f(y)\nu'(dy).
\end{aligned}$$

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:

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

where the last two inequalities follow from (A.3) and (A.4) respectively. The expected result follows and the proof of the lemma is complete. □

Proof of Theorem 4.11. 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 A.1, for all $\varepsilon > 0$, there exists a kernel p_1^ε such that

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

where $\nu, \nu_1', \nu_2, \nu_2'$ are defined in Lemma A.1. Applying the transport-entropy inequalities that hold for μ_1 and μ_2, we get

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

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 we set $a_1 := \max(a_1^{(1)}, a_1^{(2)})$ and $a_2 := \max(a_2^{(1)}, a_2^{(2)})$). Letting ε go to zero completes the proof of the theorem. □

Remark A.5. *Alternatively, following [55], one could give a proof based on the dual characterization of Proposition 4.5.*

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. N. Alon and J. H. Spencer, *The probabilistic method*, third ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons, Inc., Hoboken, NJ, 2008.
3. 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.
4. L. Ambrosio, N. Gigli, and G. 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.
5. 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.
6. S. G. Bobkov, *An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space*, Ann. Probab. **25** (1997), no. 1, 206–214.
7. S. G. Bobkov, I. Gentil, and M. Ledoux, *Hypercontractivity of Hamilton-Jacobi equations*, J. Math. Pures Appl. (9) **80** (2001), no. 7, 669–696.
8. 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.
9. S. G. Bobkov, C. Houdré, and P. Tetali, *The subgaussian constant and concentration inequalities*, Israel J. Math. **156** (2006), 255–283.
10. S. G. Bobkov and M. Ledoux, *Poincaré's inequalities and Talagrand's concentration phenomenon for the exponential distribution*, Probab. Theory Related Fields **107** (1997), no. 3, 383–400.
11. S. G. Bobkov and M. Ledoux, *From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities*, Geom. Funct. Anal. **10** (2000), no. 5, 1028–1052.
12. S. G. Bobkov and P. Tetali, *Modified logarithmic Sobolev inequalities in discrete settings*, J. Theoret. Probab. **19** (2006), no. 2, 289–336.
13. V. I. Bogachev, *Measure theory. Vol. II*, Springer-Verlag, Berlin (2007).
14. A.-I. Bonciocat and K.-T. Sturm, *Mass transportation and rough curvature bounds for discrete spaces*, J. Funct. Anal. **256** (2009), no. 9, 2944–2966.
15. C. Borell, *The Brunn-Minkowski inequality in Gauss space*, Invent. Math. **30** (1975), no. 2, 207–216.
16. S. Boucheron, G. Lugosi, and P. Massart, *Concentration inequalities*, Oxford University Press, Oxford, 2013, A nonasymptotic theory of independence, With a foreword by Michel Ledoux.
17. A. Dembo, *Information inequalities and concentration of measure*, Ann. Probab. **25** (1997), no. 2, 927–939.
18. J.-D. Deuschel and D.W. Stroock, *Large deviations*, Pure and Applied Mathematics, vol. 137, Academic Press Inc., Boston, MA, 1989.
19. H. Djellout, A. Guillin, and L. Wu, *Transportation cost-information inequalities and applications to random dynamical systems and diffusions*, Ann. Probab. **32** (2004), no. 3B, 2702–2732.
20. D. P. Dubhashi and A. Panconesi, *Concentration of measure for the analysis of randomized algorithms*, Cambridge University Press, Cambridge, 2009.
21. M. Erbar and J. Maas, *Ricci curvature of finite Markov chains via convexity of the entropy*, Arch. Ration. Mech. Anal. **206** (2012), no. 3, 997–1038.
22. P. Federbush, *A partially alternate derivation of a result of Nelson*, J. Physics **10** (1969), 50–52.
23. N. Gozlan, *A characterization of dimension free concentration in terms of transportation inequalities*, Ann. Probab. **37** (2009), no. 6, 2480–2498.

24. N. Gozlan and C. Léonard, *Transport inequalities. A survey*, Markov Process. Related Fields **16** (2010), no. 4, 635–736.
25. 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.
26. N. Gozlan, C. Roberto, and P.-M. Samson, *A new characterization of Talagrand's transport-entropy inequalities and applications*, Ann. Probab. **39** (2011), no. 3, 857–880.
27. N. Gozlan, C. Roberto, and P.-M. Samson, *Characterization of Talagrand's transport-entropy inequalities in metric spaces*, Ann. Probab. **41** (2013), no. 5, 3112–3139.
28. N. Gozlan, C. Roberto, and P.-M. Samson, *Hamilton Jacobi equations on metric spaces and transport entropy inequalities*, Rev. Mat. Iberoam. **30** (2014), no. 1, 133–163.
29. N. Gozlan, C. Roberto, P.-M. Samson, Y. Shu, and P. Tetali, *Characterization of a class of weak transport-entropy inequalities on the line*, Preprint - to appear in Ann. Inst. H. Poincaré Probab. Statist. (2017)
30. N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali, *Displacement convexity of entropy and related inequalities on graphs*, Probab. Theory Related Fields **160** (2014), no. 1-2, 47–94.
31. M. Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Progress in Mathematics, vol. 152, Birkhäuser Boston, Inc., Boston, MA, 1999.
32. M. Gromov and V. D. Milman, *A topological application of the isoperimetric inequality*, Amer. J. Math. **105** (1983), no. 4, 843–854.
33. L. Gross, *Logarithmic Sobolev inequalities*, Amer. J. Math. **97** (1975), no. 4, 1061–1083.
34. E. Hillion, *Concavity of entropy along binomial convolutions*, Electron. Commun. Probab. **17** (2012), no. 4, 9.
35. J.-B. Hiriart-Urruty and C. Lemaréchal, *Fundamentals of convex analysis*, Grundlehren Text Editions, Springer-Verlag, Berlin, 2001.
36. M. Ledoux, *On Talagrand's deviation inequalities for product measures*, ESAIM Probab. Statist. **1** (1995/97), 63–87 (electronic).
37. M. Ledoux, *The concentration of measure phenomenon*, Mathematical Surveys and Monographs, vol. 89, American Mathematical Society, Providence, RI, 2001.
38. C. Léonard, *A saddle-point approach to the Monge-Kantorovich optimal transport problem*, ESAIM Control Optim. Calc. Var. **17** (2011), no. 3, 682–704.
39. J. Lott and C. Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. of Math. (2) **169** (2009), no. 3, 903–991.
40. M. J. Luczak and C. McDiarmid, *Concentration for locally acting permutations*, Discrete Math.**265** (2003), no. 1-3, 159–171.
41. K. Marton, *A simple proof of the blowing-up lemma*, IEEE Trans. Inform. Theory **32** (1986), no. 3, 445–446.
42. K. Marton, *Bounding $\bar{d}$-distance by informational divergence: a method to prove measure concentration*, Ann. Probab. **24** (1996), no. 2, 857–866.
43. K. Marton, *A measure concentration inequality for contracting Markov chains*, Geom. Funct. Anal. **6** (1996), no. 3, 556–571.
44. B. Maurey, *Some deviation inequalities*, Geom. Funct. Anal. **1** (1991), no. 2, 188–197.
45. T. Mikami, *A simple proof of duality theorem for Monge-Kantorovich problem*, Kodai Math. J. **29** (2006), no. 1, 1–4.
46. V. D. Milman, *A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies*, Funkcional. Anal. i Priložen. **5** (1971), no. 4, 28–37.
47. Y. Ollivier, *Ricci curvature of Markov chains on metric spaces*, J. Funct. Anal. **256** (2009), no. 3, 810–864.
48. Y. Ollivier and C. Villani, *A curved Brunn-Minkowski inequality on the discrete hypercube, or: what is the Ricci curvature of the discrete hypercube?*, SIAM J. Discrete Math. **26** (2012), no. 3, 983–996.
49. 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.
50. K. R. Parthasarathy, *Probability measures on metric spaces*, Probability and Mathematical Statistics, No. 3, Academic Press Inc., New York, 1967.
51. D. Paulin, *The convex distance inequality for dependent random variables, with applications to the stochastic travelling salesman and other problems*, Electron. J. Probab. **19** (2014), no. 68, 34.
52. L. Saloff-Coste, *Lectures on finite Markov chains*, Lectures on probability theory and statistics (Saint-Flour, 1996), Lecture Notes in Math., vol. 1665, Springer, Berlin, 1997, pp. 301–413.
53. P.-M. Samson, *Concentration of measure inequalities for Markov chains and Φ-mixing processes*, Ann. Probab. **28** (2000), no. 1, 416–461.
54. P.-M. Samson, *Concentration inequalities for convex functions on product spaces*, Stochastic inequalities and applications, Progr. Probab., vol. 56, Birkhäuser, Basel, 2003, pp. 33–52.
55. P.-M. Samson, *Infimum-convolution description of concentration properties of product probability measures, with applications*, Ann. Inst. H. Poincaré Probab. Statist. **43** (2007), no. 3, 321–338.
56. P.-M. Samson, *Transport-entropy inequalities on locally acting groups of permutations*, to appear in Electron. J. Probab. (2017)
57. Y. Shu, *Hamilton Jacobi equation on graphs and applications*, In preparation, 2014.
58. A. J. Stam, *Some inequalities satisfied by the quantities of information of Fisher and Shannon*, Information and Control **2** (1959), 101–112.
59. V. Strassen, *The existence of probability measures with given marginals*, Ann. Math. Statist. **36** (1965), 423–439.
60. K.-T. Sturm, *On the geometry of metric measure spaces. I*, Acta Math. **196** (2006), no. 1, 65–131.

61. V. N. Sudakov and B. S. Cirel′son, *Extremal properties of half-spaces for spherically invariant measures*, Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) **41** (1974), 14–24, 165, Problems in the theory of probability distributions, II.
62. M. Talagrand, *Concentration of measure and isoperimetric inequalities in product spaces*, Inst. Hautes Études Sci. Publ. Math. (1995), no. 81, 73–205.
63. M. Talagrand, *New concentration inequalities in product spaces*, Invent. Math. **126** (1996), no. 3, 505–563.
64. M. Talagrand, *Transportation cost for Gaussian and other product measures*, Geom. Funct. Anal. **6** (1996), no. 3, 587–600.
65. X. Tan and N. Touzi, *Optimal transportation under controlled stochastic dynamics*, Ann. Probab. **41** (2013), no. 5, 3201–3240.
66. C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.
67. C. Villani, *Optimal transport*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338, Springer-Verlag, Berlin, 2009, Old and new.
68. 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, croberto@math.cnrs.fr, paul-marie.samson@univ-mlv.fr, tetali@math.gatech.edu