DISPLACEMENT CONVEXITY OF ENTROPY AND RELATED INEQUALITIES ON GRAPHS

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

Abstract. We introduce the notion of an interpolating path on the set of probability measures on finite graphs. Using this notion, we first prove a displacement convexity property of entropy along such a path and derive Prekopa-Leindler type inequalities, a Talagrand transport-entropy inequality, certain HWI type as well as log-Sobolev type inequalities in discrete settings. To illustrate through examples, we apply our results to the complete graph and to the hypercube for which our results are optimal – by passing to the limit, we recover the classical log-Sobolev inequality for the standard Gaussian measure with the optimal constant.

1. Introduction

In recent years, Optimal Transport and its link with the Ricci curvature in Riemannian geometry attracted a considerable amount of attention. The extensive modern book by C. Villani [57] is one of the main references on this topic. However, while a lot is now known in the Riemannian setting (and more generally in geodesic spaces), very little is known so far in discrete spaces (such as finite graphs or finite Markov chains), with the notable exception of some notions of (discrete) Ricci curvature proposed recently by several authors – unfortunately there is not yet a satisfactory (universally agreed upon) resolution even there – see Bonciocat-Sturm [6], Erbar-Maas [13], Hillion [19], Joulin [23], Lin-Yau [30], Maas [32], Mielke [38], Ollivier [39], and recent works on the displacement convexity of entropy by Hillion [20], Lehec[26] and Léonard [29].

In particular, the notions of Transport inequalities, HWI inequalities, interpolating paths on the measure space, displacement convexity of entropy, are yet to be properly introduced, analyzed and understood in discrete spaces. This is the chief aim of the present paper, and of a companion paper [17]. Due to its theoretical as well as applied appeal, this subject is at the intersection of many areas of mathematics, such as Calculus of Variations, Probability Theory, Convex Geometry and Analysis, as well as Combinatorial Optimization.

In order to present our results, let us first introduce some of the relevant notions in the continuous framework of geodesic spaces, see [57].

A complete, separable, metric space $(\mathcal{X}, d)$ is said to be a *geodesic space*, if for all $x_0, x_1 \in \mathcal{X}$, there exists at least one path $\gamma\colon [0,1] \mapsto \mathcal{X}$ such that $\gamma(0) = x_0, \gamma(1) = x_1$ and

$$d(\gamma(s), \gamma(t)) = |t - s| d(x_0, x_1), \qquad \forall s, t \in [0,1].$$

Such a path is then called a *constant speed geodesic* between x_0 and x_1.

Date: August 5, 2012.

Key words and phrases. Displacement convexity, transport inequalities, modified logarithmic-Sobolev inequalities, Ricci curvature.

Supported by the grants ANR 2011 BS01 007 01, ANR 10 LABX-58, and the NSF DMS 1101447.

Then, for $p \geq 1$, let $\mathcal{P}_p(\mathcal{X})$ be the set of Borel probability measures on $\mathcal{X}$ having a finite p-th moment, namely

$$\mathcal{P}_p(\mathcal{X}) := \left\{ \mu \text{ Borel probability measure } : \int_{\mathcal{X}} d(x_o, x)^p \mu(dx) < +\infty \right\},$$

where $x_o \in \mathcal{X}$ is arbitrary ($\mathcal{P}_p(\mathcal{X})$ does not depend on the choice of the point x_o) and define the following L_p-Wasserstein distance: for $\nu_0, \nu_1 \in \mathcal{P}_p(\mathcal{X})$, set

$$W_p(\nu_0, \nu_1) := \left(\inf_{\pi \in \Pi(\nu_0, \nu_1)} \left\{ \iint d(x,y)^p \, d\pi(x,y) \right\} \right)^{1/p}, \tag{1.1}$$

where $\Pi(\nu_0, \nu_1)$ is the set of couplings of ν_0 and ν_1.

The metric space $(\mathcal{P}_p(\mathcal{X}), W_p)$ is canonically associated to the original metric space $(\mathcal{X}, d)$. Namely, if $p > 1$, $(\mathcal{P}_p(\mathcal{X}), W_p)$ is geodesic if and only if $(\mathcal{X}, d)$ is geodesic, see [54].

A remarkable and powerful fact is that, when $\mathcal{X}$ is a Riemannian manifold, one can relate the Ricci curvature of the space to the convexity of entropy along geodesics [36, 8, 45, 31, 53, 56]. More precisely, under the Bakry-Emery CD(K, ∞) condition (see e.g. [2]), namely if the space $(\mathcal{X}, d, \mu)$ is such that Ric + Hess $V \geq K$, where $\mu(dx) = e^{-V(x)}\, dx$, then one can prove that for all $\nu_0, \nu_1 \in \mathcal{P}_2(\mathcal{X})$ whose supports are included in the support of μ, there exists a constant speed W_2-geodesic $\{\nu_t\}_{t \in [0,1]}$ from ν_0 to ν_1 such that

$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{K}{2} t(1-t) W_2^2(\nu_0, \nu_1) \qquad \forall t \in [0,1], \tag{1.2}$$

where $H(\nu|\mu)$ denotes the relative entropy of ν with respect to μ. Equation (1.2) is known as the *K-displacement convexity of entropy*. In fact, a converse statement also holds: if the entropy is K-displacement convex, then the Ricci curvature is bounded below by K. This equivalence was used as a guideline for the definition of the notion of curvature in geodesic spaces by Sturm-Lott-Villani in their celebrated works [31, 54, 55].

Moreover, it is known that the K-displacement convexity of entropy is a very strong notion that implies many well-known inequalities in Convex Geometry and in Probability Theory, such as the Brunn-Minkowski inequality, the Prekopa-Leindler inequality, Talagrand's transport-entropy inequality, HWI inequality, log-Sobolev inequality etc., see [57].

The question one would like to address is whether one can extend the above theory to discrete settings such as finite graphs, equipped with a set of probability measures on the vertices and with a natural graph distance.

Let us mention two main obstructions. Firstly, W_2-geodesics do not exist in discrete settings (the reader can verify this fact by considering two nearest neighbors x, y in the graph $G = (V, E)$ and constructing a constant speed geodesic between the two Dirac measures δ_x, δ_y at the vertices x and y). On the other hand, the following Talagrand's transport-entropy inequality

$$W_2^2(\nu_0, \mu) \leq C\, H(\nu_0|\mu), \qquad \forall \nu_0 \in \mathcal{P}_2(V) \tag{1.3}$$

(for a suitable constant $C > 0$) does not hold in discrete settings unless μ is a Dirac measure! From these simple observations we deduce that W_2 is not well adapted either for defining the path $\{\nu_t\}_{t \in [0,1]}$ or for measuring the defect/excess in the convexity of entropy in a discrete context.

In this paper, our contribution is to introduce the notion of an interpolating path $\{\nu_t\}_{t \in [0,1]}$ and of a *weak transport cost* $\widetilde{T}_2$ (that in a sense goes back to Marton [33, 34]). These will in turn help us derive the desired displacement convexity results on finite graphs.

Before presenting our results, we give a brief state of the art of the field (to the best of our knowledge).

In [40], Ollivier and Villani prove that, on the hypercube $\Omega_n = \{0,1\}^n$, for any probability measures ν_0, ν_1, there exists a probability measure $\nu_{1/2}$ (concentrated on the set of mid-points, see [40] for a precise definition) such that

$$H(\nu_{1/2}|\mu) \leq \frac{1}{2}H(\nu_0|\mu) + \frac{1}{2}H(\nu_1|\mu) - \frac{1}{80n}W_1^2(\nu_0,\nu_1)\,,$$

where $\mu \equiv 1/2^n$ is the uniform measure and W_1 is defined with the Hamming distance. They observe that, this in turn implies some curved Brunn-Minkowski inequality on Ω_n. The constant $1/n$ encodes, in some sense, the discrete Ricci curvature of the hypercube in accordance with the various definitions of the discrete Ricci curvature (see above for references).

Maas [32] introduces a *pseudo* Wasserstein distance $\mathcal{W}_2$ that corresponds to the geodesic distance on the set, $\mathcal{P}(\Omega_n)$, of probability measures on the hypercube Ω_n, equipped with a Riemannian metric. (In fact, his construction is more general and applies to a wide class of Markov kernels on finite graphs.) This metric is such that the continuous time random walk on the graph becomes a gradient flow of the function $H(\cdot|\mu)$. This is further developed by Erbar and Maas [13] who prove, inter alia, that if $\{\nu_t\}_{t\in[0,1]}$ is a geodesic from ν_0 to ν_1, then

$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{1}{n}t(1-t)\mathcal{W}_2^2(\nu_0,\nu_1)\,, \qquad \forall t \in [0,1]\,,$$

where $\mu \equiv 1/2^n$ is the uniform measure. Independently, Mielke [38] also obtains similar results. As a consequence of their displacement convexity property, these authors derive versions of log-Sobolev, HWI and Talagrand's transport-entropy inequalities (involving $\mathcal{W}_2$ and W_1 distances) with sharp constants. Very recent works of Erbar [12] and Gigli-Maas [15] derive further results with the pseudo metric, demonstrating that the metric also works, in a certain sense, in continuous settings.

In a different direction (at the level of functional inequalities), besides the study of the log-Sobolev inequality which is now classical (see *e.g.* [48, 1]), Sammer and the last named author [50, 49] studied Talagrand's inequality in discrete spaces, with W_1 on the left hand side of (1.3). They also derived a discrete analogue of the Otto-Villani result [41]: that a modified log-Sobolev inequality implies the W_1-type Talagrand inequality. Connected to this, a few years ago, following seminal work of Bobkov and Ledoux [3], several researchers independently realized that modified versions of logarithmic Sobolev inequalities helped capture refined information that was lost while working with the classic log-Sobolev inequality of Gross. In the discrete setting of finite Markov chains, one such modified log-Sobolev inequality has been instrumental in capturing the rate of convergence to equilibrium in the (relative) entropy sense, see e.g. [7], [10], [5], [14], [16], [48], [46]. The current state of knowledge in identifying precise sufficient criteria to derive bounds on the entropy decay (or on the corresponding modified log-Sobolev constants) is unfortunately rather meagre. This is an independent motivation for our efforts at developing the discrete aspects of the displacement convexity property and related notions.

Now we describe some of the main results of the present paper. At first, we shall introduce the notion of an interpolating path $\{\nu_t^\pi\}_{t\in[0,1]}$, on the set of probability measures on graphs, between two arbitrary probability measures ν_0, ν_1. In fact, we define a *family* of interpolating paths, depending on a parameter $\pi \in \Pi(\nu_0,\nu_1)$, which is a coupling of ν_0, ν_1. The construction of this interpolating path is inspired by a certain binomial interpolation due to Johnson [22], see also [19, 20, 21]. In particular, we shall prove that such an interpolating path, for a properly chosen coupling π^* – namely an optimal coupling for W_1 – is actually a W_1 constant speed geodesic: *i.e.*, $W_1(\nu_t^{\pi^*}, \nu_s^{\pi^*}) = |t-s|W_1(\nu_0,\nu_1)$ for all $s,t \in [0,1]$, with W_1 defined using the graph distance d (see Proposition 2.5 below). Such a family enjoys a tensorisation (see Lemma 2.10) that is crucial in our derivation of the displacement convexity property on product of graphs.

Indeed, we shall prove the following tensoring property of a displacement convexity of entropy along the interpolating path $\{\nu_t^\pi\}_{t\in[0,1]}$. This is one of our main results (see below and Theorem 4.6). In order to state the result, we define here the notion of a quadratic cost, which we will elaborate on, in the later sections.

Let $G=(V,E)$ be a (finite) connected, undirected graph, and let $\mathcal{P}(V)$ denote the set of probability measures on the vertex set V. Given two probability measures ν_0 and ν_1 on V, let $\Pi(\nu_0,\nu_1)$ denote the set of couplings (joint distributions) of ν_0 and ν_1. Given $\pi \in \Pi(\nu_0,\nu_1)$, consider the probability kernels p and $\bar{p}$ defined by

$$\pi(x,y) = \nu_0(x)p(x,y) = \nu_1(y)\bar{p}(y,x), \quad \forall x,y \in V,$$

and set

$$I_2(\pi) := \sum_{x\in V}\left(\sum_{y\in V} d(x,y)p(x,y)\right)^2 \nu_0(x), \tag{1.4}$$

$$\bar{I}_2(\pi) := \sum_{y\in V}\left(\sum_{x\in V} d(x,y)\bar{p}(y,x)\right)^2 \nu_1(y)\,.$$

We say a graph G, equipped with the distance d and probability measure $\mu \in \mathcal{P}(V)$, satisfies the displacement convexity property (of entropy), if there exists a $C = C(G,d,\mu) > 0$, so that for any $\nu_0,\nu_1 \in \mathcal{P}(V)$, there exists a $\pi \in \Pi(\nu_0,\nu_1)$ satisfying:

$$H(\nu_t^\pi|\mu) \le (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - Ct(1-t)(I_2(\pi)+\bar{I}_2(\pi))\,, \qquad \forall t\in[0,1].$$

The quantity $I_2(\pi)$ goes back to Marton [33, 34] in her definition of the following transport cost, we call *weak transport cost*:

$$\widetilde{W}_2^2(\nu_0,\nu_1) := \inf_{\pi\in\Pi(\nu_0,\nu_1)} I_2(\pi) + \inf_{\pi\in\Pi(\nu_0,\nu_1)} \bar{I}_2(\pi)\,.$$

For more on this Wasserstein-type distance, see [11, 35, 51]. The precise statement of our tensorisation theorem is as follows. For a graph, by the *graph distance* between two vertices, we mean the length of a shortest path between the two vertices.

Theorem 1.5. *For $i \in \{1,\dots,n\}$, let μ^i be a probability measure on $G_i = (V_i,E_i)$, with the graph distance d_i. Assume also that for each $i \in \{1,\dots,n\}$ there is a constant $C_i \ge 0$ such that for all probability measures ν_0,ν_1 on V_i, there exists $\pi = \pi^i \in \Pi(\nu_0,\nu_1)$ such that it holds*

$$H(\nu_t^\pi|\mu^i) \le (1-t)H(\nu_0|\mu^i) + tH(\nu_1|\mu^i) - C_it(1-t)(I_2(\pi)+\bar{I}_2(\pi)) \qquad \forall t\in[0,1].$$

Then the product probability measure $\mu = \mu^1 \otimes \cdots \otimes \mu^n$ defined on the Cartesian product $G = G_1 \square \cdots \square G_n$ (see below for a precise definition) verifies the following property: for all probability measures ν_0,ν_1 on V, there exists $\pi = \pi^{(n)} \in \Pi(\nu_0,\nu_1)$ satisfying,

$$H(\nu_t^\pi|\mu) \le (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - Ct(1-t)(I_2^{(n)}(\pi)+\bar{I}_2^{(n)}(\pi)) \qquad \forall t\in[0,1],$$

where $C = \min_i C_i$*,*

$$I_2^{(n)}(\pi) := \sum_{x \in V_1 \times \cdots \times V_n} \sum_{i=1}^n \left(\sum_{y \in V_1 \times \cdots \times V_n} d_i(x_i, y_i) \frac{\pi(x,y)}{\nu_0(x)} \right)^2 \nu_0(x),$$

and

$$\bar{I}_2^{(n)}(\pi) := \sum_{y \in V_1 \times \cdots \times V_n} \sum_{i=1}^n \left(\sum_{x \in V_1 \times \cdots \times V_n} d_i(x_i, y_i) \frac{\pi(x,y)}{\nu_1(y)} \right)^2 \nu_1(y).$$

(and with $I_2(\pi) := I_2^{(1)}(\pi)$ *and similarly for* $\bar{I}_2(\pi)$*).*

In particular, as a consequence of the above tensorisation theorem, we shall prove that, given two probability measures ν_0, ν_1 on the hypercube $\Omega_n = \{0,1\}^n$, there exists a coupling π such that

$$H(\nu_t^\pi|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{1}{2}t(1-t)\widetilde{W}_2^2(\nu_0, \nu_1), \qquad \forall t \in [0,1] \tag{1.6}$$

where $\mu \equiv 1/2^n$ is the uniform measure (but that could be any product of Bernoulli measures). As it is easy to see, the weak transport cost $\tilde{W}_2$ is weaker than W_2, but stronger than W_1. Moreover, $\widetilde{W}_2^2(\nu_0, \nu_1) \geq \frac{2}{n} W_1^2(\nu_0, \nu_1)$ (see below) so that (1.6) captures, in a sense, a discrete Ricci curvature of the hypercube (see [40] and references therein).

As a by-product of the displacement convexity property above, we shall derive a series of consequences. More precisely, we shall first derive a so-called HWI inequality.

Proposition 1.7. *Let* μ *be a probability measure on* V^n*, the vertex set of a product graph. Assume that* μ *verifies the following displacement convexity inequality: there is some* $c > 0$ *such that for any probability measures* ν_0, ν_1 *on* V^n*, there exists a coupling* $\pi \in \Pi(\nu_0, \nu_1)$ *such that*

$$H(\nu_t^\pi|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - ct(1-t)(I_2^{(n)}(\pi) + \bar{I}_2^{(n)}(\pi)) \qquad \forall t \in [0,1].$$

Then μ *verifies*

$$H(\nu_0|\mu) \leq H(\nu_1|\mu) + \sqrt{\sum_{x \in V^n} \sum_{i=1}^n \left[\sum_{z \in N_i(x)} \left(\log \frac{\nu_0(x)}{\mu(x)} - \log \frac{\nu_0(z)}{\mu(z)} \right) \right]_+^2 \nu_0(x)} \sqrt{I_2^{(n)}(\pi)} - c(I_2^{(n)}(\pi) + \bar{I}_2^{(n)}(\pi)),$$

for the same $\pi \in \Pi(\nu_0, \nu_1)$ *as above, where* $N_i(x)$ *is the set of neighbors of* x *in the* i*-th direction (see Proposition 5.1 for a precise definition).*

On the hypercube, the latter implies the following log-Sobolev-type inequality (that can be seen as a reinforcement of a discrete modified log-Sobolev inequality (see Corollary 5.3)): if $\mu \equiv 1/2^n$, for any $f \colon \Omega_n \to (0, \infty)$, it holds

$$\mathrm{Ent}_\mu(f) \leq \frac{1}{2} \sum_{x \in \Omega_n} \sum_{i=1}^n [\log f(x) - \log f(\sigma_i(x))]_+^2 f(x)\mu(x) - \frac{1}{2}\widetilde{W}_2^2(f\mu|\mu),$$

where $\sigma_i(x) = (x_1, \ldots, x_{i-1}, 1 - x_i, x_{i+1}, \ldots, x_n)$ is the vector $x = (x_1, \ldots, x_n)$ with the i-th coordinate flipped, and the constant $1/2$ (in front of the Dirichlet form) is optimal.

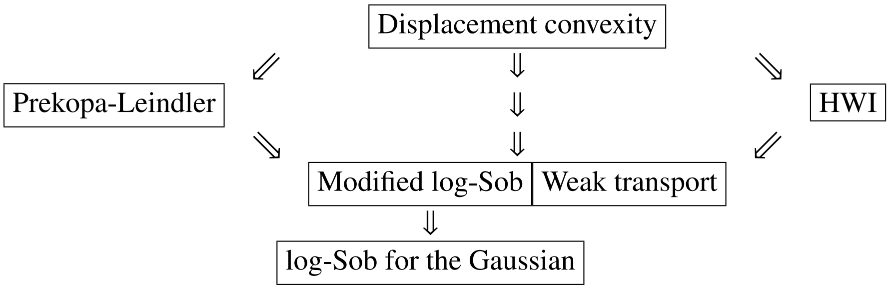
From this, by means of the Central Limit Theorem, the above reinforced modified log-Sobolev inequality actually leads to the usual logarithmic Sobolev inequality of Gross [18] for the standard Gaussian, with the optimal constant (see Corollary 5.5).

In a different direction, we also prove that the displacement convexity along the interpolating path $\{\nu_t^\pi\}_{t\in[0,1]}$ implies a discrete Prekopa-Leindler Inequality (Theorem 6.4), which in turn, as in the continuous setting, implies a logarithmic Sobolev inequality and a (weak) transport-entropy inequality of the Talagrand-type:

$$\widetilde{W}_2^2(\nu|\mu) \leq C\, H(\nu|\mu)\,, \qquad \forall \nu$$

for a suitable constant $C > 0$. These implications and inequalities are studied in further detail – their various links with the concentration of measure phenomenon and with other functional inequalities – in the companion paper [17].

We may summarize the various implications that we prove in the following diagram:



In summary, our paper develops various theoretical objects of much current interest (the interpolating path $\{\nu_t^\pi\}_{t\in[0,1]}$, the weak transport cost $\widetilde{W}_2$, the displacement convexity property and its consequences) in a *discrete* context. Our concrete examples include the complete graph and the hypercube. However, our theory applies to other graphs (not necessarily product type) that we will collect in a forthcoming paper. Also, we believe that our results open a wide class of new problems and new directions of investigation in Probability Theory, Convex Geometry and Analysis.

Finally, we mention that, during the final preparation of this work, we learned that Erwan Hillion independently introduced the same kind of interpolating path, but between a Dirac at a fixed point $o \in G$ of the graph and any arbitrary measure (hence without coupling π), and derive a certain displacement convexity property [20] along the interpolation. In [20], the author also deals with the $f \cdot g$ decomposition introduced by Léonard [29].

Our presentation follows the following table of contents.

CONTENTS

1.1. **Notation.** Throughout the paper we shall use the following notation.

Graphs. $G = (V, E)$ will denote a finite connected undirected graph with the vertex set V and the edge set E. For any two vertices x and y of G, $x \sim y$ means that x and y are nearest neighbors (for the graph structure of G), *i.e.* $(x, y) \in E$. We use d for the graph distance defined below.

Given two graphs $G_1 = (V_1, E_1)$, $G_2 = (V_2, E_2)$, with graph distances d_1, d_2 respectively, we set $G_1 \square G_2 = (V_1 \times V_2, E_1 \square E_2)$ for the Cartesian product of the two graphs, equipped with the ℓ^1 distance $d(x, y) = d_1(x_1, y_1) + d_2(x_2, y_2)$, for all $x = (x_1, x_2), y = (y_1, y_2) \in G_1 \times G_2$. More precisely, $((x_1, x_2), (y_1, y_2)) \in E_1 \square E_2$ if either $x_1 = y_1$ and $x_2 \sim y_2$, or $x_1 \sim y_1$ and $x_2 = y_2$. The Cartesian product of G with itself will simply be denoted by G^2, and more generally by G^n, for all $n \geq 2$.

Paths and geodesics. A *path* $\gamma = (x_0, x_1, \dots, x_n)$ (of G) is an oriented sequence of vertices of G satisfying $x_{i-1} \sim x_i$ for any $i = 1 \dots, n$. Such a path starts at x_0 and ends at x_n and is said to be of length $|\gamma| = n$. The graph distance $d(x, y)$ between two vertices $x, y \in G$ is the minimal length of a path connecting x to y. Any path of length $n = d(x, y)$ between x and y is called a *geodesic* between x and y. By construction, any geodesic is self-avoiding. We will denote by $\Gamma(x, y)$ the set of all geodesics from x to y.

We will say that a path $\gamma = (x_0, x_1, \dots, x_n)$ crosses the vertex $z \in V$, if there is some k such that $z = x_k$. In this case, we will write $z \in \gamma$. Given $z \in V$, we set $C(z) = \{(x, y) \text{ such that } z \in \gamma \text{ for some } \gamma \in \Gamma(x, y)\}$ for the set of couples such that some geodesic joining them goes through z. Conversely, if z belongs to some geodesic between x and y, we shall write $z \in [\![x, y]\!]$ and say that z is *between* x and y. Finally, for all $x, y, z \in V$, we will denote by $\Gamma(x, z, y)$, the set of geodesics $\gamma \in \Gamma(x, y)$ such that $z \in \gamma$. This set is nonempty if and only if $z \in [\![x, y]\!]$.

Probability measures and couplings. We write $\mathcal{P}(V)$ for the set of probability measures on V. Given a probability measure $\nu \in \mathcal{P}(V)$ and a function $f\colon V \to \mathbb{R}$, $\nu(f) = \sum_{z \in V} \nu(z) f(z)$ denotes the mean value of f with respect to ν. We may also use the alternative notation $\nu(f) = \int f(x)\, \nu(dx) = \int f(x)\, d\nu(x) = \int f\, d\nu$.

Let $\nu, \mu \in \mathcal{P}(V)$; the *relative entropy* of ν with respect to μ is defined by

$$H(\nu|\mu) = \begin{cases} \int \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu}\, d\mu & \text{if } \nu \ll \mu \\ +\infty & \text{otherwise} \end{cases}$$

where $\nu \ll \mu$ means that ν is absolutely continuous with respect to μ, and $\frac{d\nu}{d\mu}$ denotes the density of ν with respect to μ.

Given a density $f\colon V \to (0,\infty)$ with respect to a given probability measure μ (*i.e.* $\mu(f) = 1$), we shall use the following notation for the relative entropy of $f\mu$ with respect to μ:

$$\mathrm{Ent}_\mu(f) := H(f\mu|\mu) = \int f \log f d\mu.$$

If $f\colon V \to (0,\infty)$ is no longer a density, then $\mathrm{Ent}_\mu(f) := \int f \log(f/\mu(f))\, d\mu$.

Given two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ and a probability measure $\mu \in \mathcal{P}(V_1 \times V_2)$ on the product, we *disintegrate* μ as follows: let μ^2 be the second marginal of μ, *i.e.* $\mu^2(x_2) = \sum_{x_1\in V_1} \mu(x_1, x_2) = \mu(V_1, x_2)$, for all $x_2 \in V_2$, and set $\mu^1(x_1|x_2)$ so that

$$\mu(x_1, x_2) = \mu^2(x_2)\mu^1(x_1|x_2), \qquad \forall (x_1, x_2) \in V_1 \times V_2, \tag{1.8}$$

with the convention that $\mu^1(x_1|x_2) = 0$ if $\mu^2(x_2) = 0$. Equation (1.8) will be referred to as the *disintegration formula* of μ.

Recall that a coupling π of two probability measures μ and ν in $\mathcal{P}(V)$ is a probability measure on V^2 so that μ and ν are its first and second marginals, respectively: *i.e.* $\pi(x, V) = \mu(x)$ and $\pi(V, y) = \nu(y)$, for all $x, y \in V$. Given $\mu, \nu \in \mathcal{P}(V)$, the set of all couplings of μ and ν will be denoted by $\Pi(\mu, \nu)$.

Moreover, given two probability measures μ and ν in $\mathcal{P}(V)$, we denote by $P(\mu, \nu)$ the set of probability kernels[1] p such that

$$\sum_{x\in V} \mu(x)p(x, y) = \nu(y)\,, \qquad \forall y \in V.$$

By construction, given $p \in P(\mu, \nu)$, one defines a coupling $\pi \in \Pi(\mu, \nu)$ by setting $\pi(x, y) = \mu(x)p(x, y)$, $x, y \in V$. Conversely, given a coupling $\pi \in \Pi(\mu, \nu)$, we canonically construct a kernel $p \in P(\mu, \nu)$ by setting $p(x, y) = \pi(x, y)/\mu(x)$ when $\mu(x) \neq 0$ and $p(x, y) = 0$ otherwise.

Warning 1: In the sequel, it will always be understood, although not explicitly stated, that $p(x, y) = 0$ if $\mu(x) = 0$ and similarly in the disintegration formula (1.8).

Warning 2: Throughout, we will use the French notation $C_n^k := \binom{n}{k} = \frac{n!}{k!(n-k)!}$ for the binomial coefficients.

2. A NOTION OF A PATH ON THE SET OF PROBABILITY MEASURES ON GRAPHS.

The aim of this section is to define a class of paths between probability measures on graphs. As proved below, each path in this class is a geodesic, in the space of probability measures equipped with the Wasserstein distance W_1 (see below). It satisfies a convenient differentiation property and also has the nice feature of allowing tensorisation. We shall end the section with some specific examples.

2.1. Construction.

Inspired by [22], we will first construct an interpolating path between two Dirac measures δ_x and δ_y, for arbitrary $x, y \in V$, on the set of probability measures $\mathcal{P}(V)$. Fix $x, y \in V$ and denote by Γ the random variable that chooses uniformly at random a geodesic γ in $\Gamma(x, y)$. Also, for any $t \in [0, 1]$, let $N_t \sim \mathcal{B}(d(x, y), t)$ be a binomial variable of parameter $d(x, y)$ and t, independent of Γ (observe that $N_0 = 0$ and $N_1 = d(x, y)$). Then denote by $X_t = \Gamma_{N_t}$ the random position on Γ after N_t jumps starting from x. Finally, set $\nu_t^{x,y}$ for the law of X_t.

---

[1] We recall that $p : V \times V \to [0, 1]$ is a probability kernel if, for all $x \in V$, $\sum_{y\in V} p(x, y) = 1$.

By construction, $\nu_t^{x,y}$ is clearly a path from δ_x to δ_y. Moreover, for all $z \in V$, we have

$$\nu_t^{x,y}(z) = \sum_{\gamma \in \Gamma(x,y)} \mathbb{P}(X_t = z | \Gamma = \gamma, z \in \Gamma)\mathbb{P}(\Gamma = \gamma, z \in \gamma) = \sum_{\gamma \in \Gamma(x,y)} C_{d(x,y)}^{d(x,z)} t^{d(x,z)}(1-t)^{d(y,z)} \frac{\mathbb{1}_{z \in \gamma}}{|\Gamma(x,y)|}.$$

Therefore

$$\nu_t^{x,y}(z) = C_{d(x,y)}^{d(x,z)} t^{d(x,z)}(1-t)^{d(y,z)} \frac{|\Gamma(x,z,y)|}{|\Gamma(x,y)|}.$$

For all z between x and y we observe that

$$|\Gamma(x,z,y)| = |\Gamma(x,z)| \times |\Gamma(z,y)|, \tag{2.1}$$

since there is a one-to-one correspondence between the sets of geodesics from x to z and from z to y, and the set of geodesics from x to y that cross the vertex z, just by gluing the path from x to z to the path from z to y, and by using that $d(x,y) = d(x,z) + d(z,y)$. Therefore $\nu_t^{x,y}$ takes the form

$$\nu_t^{x,y}(z) = C_{d(x,y)}^{d(x,z)} t^{d(x,z)}(1-t)^{d(y,z)} \frac{|\Gamma(x,z)| \times |\Gamma(z,y)|}{|\Gamma(x,y)|} \mathbb{1}_{z \in [\![x,y]\!]}. \tag{2.2}$$

Observe that, for any $x, y \in V$ and any $t \in (0,1)$, $\nu_t^{x,y} = \nu_{1-t}^{y,x}$.

Remark 2.3. *In the construction above of the interpolation $\nu_t^{x,y}$, the choice of the binomial random variable for the number N_t of jumps might seem somewhat ad hoc; however, in Proposition 2.12 below, we show that in fact the choice is* necessary *for $\nu_t^{x,y}$ to tensorise over a (Cartesian) product of graphs.*

Given the family $\{\nu_t^{x,y}\}_{x,y}$, we can now construct a path from any measure $\nu_0 \in \mathcal{P}(V)$ to any measure $\nu_1 \in \mathcal{P}(V)$. Namely, given a coupling $\pi \in \mathcal{P}(V \times V)$ of ν_0 and ν_1, we define

$$\nu_t^\pi(\,\cdot\,) = \sum_{(x,y) \in V^2} \pi(x,y)\nu_t^{x,y}(\,\cdot\,), \qquad \forall t \in [0,1]. \tag{2.4}$$

By construction we have $\nu_0^\pi = \nu_0$ and $\nu_1^\pi = \nu_1$. Furthermore, observe that, if $\nu_0 = \delta_x$ and $\nu_1 = \delta_y$, then necessarily $\pi = \delta_x \otimes \delta_y$ and thus $\nu_t^\pi = \nu_t^{x,y}$.

2.2. **Geodesics for W_1.** Next we prove that, when π is well chosen, $(\nu_t^\pi)_{t \in [0,1]}$ is a geodesic from ν_0 to ν_1 on the set of probability measures $\mathcal{P}(V)$ equipped with the Wasserstein L_1-distance W_1.

Given two probability measures μ and ν on $\mathcal{P}(V)$, recall that

$$W_1(\mu,\nu) = \inf_{\pi \in \Pi(\nu_0,\nu_1)} \iint d(x,y)\,\pi(dx\,dy) = \inf_{X \sim \mu, Y \sim \nu} \mathbb{E}[d(X,Y)]$$

The following result asserts that $(\nu_t^\pi)_{t \in [0,1]}$ is actually a geodesic for W_1 when π is an optimal coupling.

Proposition 2.5. *For any probability measures $\nu_0, \nu_1 \in \mathcal{P}(V)$, it holds*

$$W_1(\nu_s^{\pi^*}, \nu_t^{\pi^*}) = |t-s| W_1(\nu_0,\nu_1) \qquad \forall s,t \in [0,1]$$

where π^* is an optimal coupling in the definition of $W_1(\nu_0,\nu_1)$ and where $\nu_t^{\pi^*}$ is defined in (2.4).

Proof. Fix two probability measures ν_0, $\nu_1 \in \mathcal{P}(V)$ and π^* an optimal coupling in the definition of $W_1(\nu_0,\nu_1)$ (since $\mathcal{P}(V)$ is compact π^* is well defined). For brevity, set $\nu_t := \nu_t^{\pi^*}$.

First, we claim that it is enough to prove that

$$W_1(\nu_s,\nu_t) \le (t-s) W_1(\nu_0,\nu_1), \qquad \forall s,t \in [0,1] \text{ with } s \le t. \tag{2.6}$$

Indeed, assume (2.6), then recalling that W_1 is a distance (see e.g. [57]), by the triangle inequality we have

$$\begin{aligned} W_1(\nu_0,\nu_1) &\le W_1(\nu_0,\nu_s)+W_1(\nu_s,\nu_t)+W_1(\nu_t,\nu_1) \le sW_1(\nu_0,\nu_1)+(t-s)W_1(\nu_0,\nu_1)+tW_1(\nu_0,\nu_1) \\ &\le W_1(\nu_0,\nu_1). \end{aligned}$$

Hence, all the inequalities used above are actually equalities, which guarantees the conclusion of the proposition and hence the claim.

Now, we prove (2.6). Let (X,Y) be a random couple of law π^*. Fix $s\le t$, it suffises to construct a random couple (X_s,X_t) with marginal laws ν_s and ν_t so that

$$\mathbb{E}[d(X_s,X_t)] \le (t-s)\mathbb{E}[d(X,Y)] = (t-s)W_1(\nu_0,\nu_1).$$

From the last observation, let us remark that such a couple (X_s,X_t) will therefore realized

$$\mathbb{E}[d(X_s,X_t)] = W_1(\nu_s,\nu_t).$$

Let $\left((U_s^i,V_t^i)\right)_{i\ge1}$ be an independent identically distributed sequence of random couples in $\{0,1\}^2$, independent of X and Y. We chose the law of (U_s^1,V_t^1) given by

$$\mathbb{P}((U_s^1,V_t^1)=(0,0))=1-s, \quad \mathbb{P}((U_s^1,V_t^1)=(0,1))=0,$$

$$\mathbb{P}((U_s^1,V_t^1)=(1,0))=t-s, \quad \mathbb{P}((U_s^1,V_t^1)=(1,1))=t,$$

so that U_s^1 and V_t^1 are Bernoulli random variables with respective parameters s and t, and we have

$$\mathbb{E}(|U_s^1-V_t^1|)=(t-s).$$

Given $(X,Y)=(x,y)$, with $x,y\in V$, let (N_s,N_t) denote the random couple defined by

$$N_s=\sum_{i=1}^{d(x,y)} U_s^i, \quad N_t=\sum_{i=1}^{d(x,y)} V_s^i.$$

Then the laws of N_s and N_t given $(X,Y)=(x,y)$ are respectively $\mathcal{B}(d(x,y),s)$ and $\mathcal{B}(d(x,y),t)$, the binomial distribution with parameters $d(x,y)$, s and t respectively.

Finally, given $(X,Y)=(x,y)$, with $x,y\in V$, let Γ denote a random geodesic chosen uniformly in $\Gamma(x,y)$, independently of the sequence $\left((U_s^i,V_t^i)\right)_{i\ge1}$, and let $X_s=\Gamma_{N_s}$ be the random position on Γ after N_s jumps and $X_t=\Gamma_{N_t}$ be the random position on Γ after N_t jumps. By definition, the law of X_s and X_t are respectively ν_s and ν_t and one has $d(X_s,X_t)=|N_s-N_t|$. Moreover, according to this construction, one has

$$\begin{aligned} \mathbb{E}[d(X_s,X_t)] = \mathbb{E}\left[|N_s-N_t|\right] &= \mathbb{E}\left[\left|\sum_{i=1}^{d(X,Y)} U_s^i - \sum_{i=1}^{d(X,Y)} V_t^i\right|\right] \\ &\le \mathbb{E}\left[\sum_{i=1}^{d(X,Y)} \left|U_s^i-V_t^i\right|\right] = \mathbb{E}\left[\sum_{i=1}^{d(X,Y)} \mathbb{E}\left[\left|U_s^i-V_t^i\right|\right]\right] = (t-s)\mathbb{E}[d(X,Y)]. \end{aligned}$$

This completes the proof of (2.6) and Proposition 2.5. □

2.3. **Differentiation property.** A second property of the path defined in (2.2) and (2.4) is the following time differentiation property.

For any z on a given geodesic γ from x to y, if $z \neq y$, let $\gamma_+(z)$ denotes the (unique) vertex on γ at distance $d(z,y)-1$ from y (and thus at distance $d(x,z)+1$ from x), and similarly if $z \neq x$, let $\gamma_-(z)$ denote the vertex on γ at distance $d(z,y)+1$ from y (and hence at distance $d(x,z)-1$ from x). In other words, following the geodesic γ from x toward y, $\gamma_-(z)$ is the vertex just anterior to z, and $\gamma_+(z)$ the vertex posterior to z.

For any real function f on V, we also define two related notions of gradient along γ: for all $z \in \gamma$, $z \neq y$,

$$\nabla_\gamma^+ f(z) = f(\gamma_+(z)) - f(z),$$

and for all $z \in \gamma$, $z \neq x$,

$$\nabla_\gamma^- f(z) = f(z) - f(\gamma_-(z)).$$

By convention, we put $\nabla_\gamma^- f(x) = \nabla_\gamma^+ f(y) = 0$, and $\nabla_\gamma^+ f(z) = \nabla_\gamma^- f(z) = 0$, if $z \notin \gamma$. Let $\nabla_\gamma f$ denote the following convex combination of these two gradients:

$$\nabla_\gamma f(z) = \frac{d(y,z)}{d(x,y)} \nabla_\gamma^+ f(z) + \frac{d(x,z)}{d(x,y)} \nabla_\gamma^- f(z).$$

Observe that, although not explicitly stated, ∇_γ depends on x and y. Finally, for all $z \in [\![x,y]\!]$, we define

$$\nabla_{x,y} f(z) = \frac{1}{|\Gamma(x,z,y)|} \sum_{\gamma \in \Gamma(x,z,y)} \nabla_\gamma f(z),$$

and when $z \notin [\![x,y]\!]$, we set $\nabla_{x,y} f(z) = 0$.

Proposition 2.7. *For all function $f \colon V \to \mathbb{R}$ and all $x, y \in V$, it holds*

$$\frac{\partial}{\partial t} \nu_t^{x,y}(f) = d(x,y) \nu_t^{x,y}(\nabla_{x,y} f).$$

As a direct consequence of the above differentiation property, we are able to give an explicit expression of the derivative (with respect to time) of the relative entropy of ν_t^π with respect to an arbitrary reference measure.

Corollary 2.8. *Let ν_0, ν_1 and μ be three probability measures on V. Assume that ν_0, ν_1 are absolutely continuous with respect to μ. Then, for any coupling $\pi \in \Pi(\nu_0, \nu_1)$, it holds*

$$\frac{\partial}{\partial t} H(\nu_t^\pi|\mu)_{|t=0} = \sum_{\substack{x,z \in V: \\ z \sim x}} \left(\log \frac{\nu_0(z)}{\mu(z)} - \log \frac{\nu_0(x)}{\mu(x)} \right) \sum_{y \in V} d(x,y) \frac{|\Gamma(x,z,y)|}{|\Gamma(x,y)|} \pi(x,y).$$

The proof of Corollary 2.8 can be found below, while some example applications will be given in the next subsection. In order to prove Proposition 2.7, we need some preparation. Recall that $\mathcal{B}(n,t)$ denotes a binomial variable of parameter n and t, and that, for any function $h \colon \{0,1,\ldots,n\} \to \mathbb{R}$, $\mathcal{B}(n,t)(h) = \sum_{k=0}^n h(k) C_n^k t^k (1-t)^{n-k}$.

Lemma 2.9. *Let $n \in \mathbb{N}^*$ and $t \in [0,1]$. For any function $h \colon \{0,1,\ldots,n\} \to \mathbb{R}$ it holds*

$$\frac{\partial}{\partial t} \mathcal{B}(n,t)(h) = \sum_{k=0}^n [(h(k+1) - h(k))(n-k) + (h(k) - h(k-1))k] \; C_n^k t^k (1-t)^{n-k},$$

with the convention that $h(-1) = h(n+1) = 0$.

Proof of Lemma 2.9. By differentiating in t, we have

$$\frac{\partial}{\partial t}\mathcal{B}(n,t)(h)=\sum_{k=0}^{n}h(k)kC_n^kt^{k-1}(1-t)^{n-k}-\sum_{k=0}^{n}h(k)(n-k)C_n^kt^k(1-t)^{n-k-1}.$$

Now, using that $1=t+(1-t)$ and that $kC_n^k=(n-k+1)C_n^{k-1}$, we get

$$kC_n^kt^{k-1}(1-t)^{n-k}=kC_n^kt^k(1-t)^{n-k}+(n-k+1)C_n^{k-1}t^{k-1}(1-t)^{n-k+1},$$

with the convention that $C_n^{-1}=0$. Similarly, using that $(n-k)C_n^k=(k+1)C_n^{k+1}$, we have

$$(n-k)C_n^kt^k(1-t)^{n-k-1}=(n-k)C_n^kt^k(1-t)^{n-k}+(k+1)C_n^{k+1}t^{k+1}(1-t)^{n-k-1}.$$

Hence,

$$\begin{aligned}\frac{\partial}{\partial t}\mathcal{B}(n,t)(h)&=\sum_{k=0}^{n}h(k)(n-k+1)C_n^{k-1}t^{k-1}(1-t)^{n-k+1}-\sum_{k=0}^{n}h(k)(n-k)C_n^kt^k(1-t)^{n-k}\\&\quad+\sum_{k=0}^{n}h(k)kC_n^kt^k(1-t)^{n-k}-\sum_{k=0}^{n}h(k)(k+1)C_n^{k+1}t^{k+1}(1-t)^{n-k-1}\\&=\sum_{k=0}^{n}(h(k+1)-h(k))(n-k)C_n^kt^k(1-t)^{n-k}+\sum_{k=0}^{n}(h(k)-h(k-1))kC_n^kt^k(1-t)^{n-k},\end{aligned}$$

with the convention that $h(-1)=h(n+1)=0$. □

We were informed by E. Hillion that the above elementary lemma also appears in his thesis [19]. We are now in a position to prove Proposition 2.7.

Proof of Proposition 2.7. Set $n=d(x,y)$ and let Γ be a random variable uniformly distributed on $\Gamma(x,y)$ and N_t be a random variable with Binomial law $\mathcal{B}(n,t)$ independent of Γ. By definition $\nu_t^{x,y}$ is the law of $X_t=\Gamma_{N_t}$. Using the independence, we have

$$\nu_t^{x,y}(f)=\mathbb{E}\left[f(X_t)\right]=\sum_{k=0}^{n}h(k)C_n^kt^k(1-t)^{n-k},$$

with $h(k)=\mathbb{E}[f(\Gamma_k)]$, $k=0,1\ldots,n$. According to Lemma 2.9, we thus get

$$\begin{aligned}\frac{\partial}{\partial t}\nu_t^{x,y}(f)&=\sum_{k=0}^{n}\left[(h(k+1)-h(k))(n-k)+(h(k)-h(k-1))k\right]\,C_n^kt^k(1-t)^{n-k}\\&=\mathbb{E}\left[(h(N_t+1)-h(N_t))(n-N_t)+(h(N_t)-h(N_t-1))N_t\right]\\&=\mathbb{E}\left[(f(\Gamma_{N_t+1})-f(\Gamma_{N_t}))d(\Gamma_{N_t},y)+(f(\Gamma_{N_t})-f(\Gamma_{N_t-1}))d(x,\Gamma_{N_t})\right]\\&=\mathbb{E}\left[(f(\Gamma^+(X_t))-f(X_t))d(X_t,y)+(f(X_t)-f(\Gamma^-(X_t)))d(x,X_t)\right]\\&=\mathbb{E}\left[d(x,y)\nabla_\Gamma f(X_t)\right].\end{aligned}$$

Finally, observe that the law of Γ knowing $X_t=z\in[\![x,y]\!]$ is uniform on $\Gamma(x,z,y)$. Indeed,

$$\mathbb{P}(\Gamma=\gamma,\ X_t=z)=\mathbb{P}(\Gamma=\gamma,\ \gamma_{N_t}=z)=\mathbb{P}(\Gamma=\gamma,\ N_t=d(x,z),\ z\in\gamma)=\frac{\mathbb{1}_{\Gamma(x,z,y)}(\gamma)}{|\Gamma(x,y)|}\mathbb{P}(N_t=d(x,z)).$$

On the other hand,

$$\mathbb{P}(X_t=z)=\nu_t^{x,y}(z)=\mathbb{P}(N_t=d(x,z))\frac{|\Gamma(x,z,y)|}{|\Gamma(x,y)|},$$

which proves the claim. By the definition of $\nabla_{x,y}f$, it thus follows that

$$\frac{\partial}{\partial t}\nu_t^{x,y}(f) = d(x,y)\,\nu_t^{x,y}(\nabla_{x,y}f),$$

which completes the proof. □

Proof of Corollary 2.8. For simplicity, let $F = \log(\nu_0/\mu)$. Observe that, since ν_0 and ν_1 are absolutely continuous with respect to μ, so is ν_t^π. Now we observe that, since $\sum_{z\in V}\frac{\partial}{\partial t}\nu_t^\pi(z) = 0$, by Proposition 2.7 (recall that $\nu_0^\pi = \nu_0$ and $\nu_0^{x,y} = \delta_x$ by construction),

$$\begin{aligned}\frac{\partial}{\partial t}H(\nu_t^\pi|\mu)_{|t=0} &= \frac{\partial}{\partial t}\left(\sum_{z\in V}\nu_t^\pi(z)\log\frac{\nu_t^\pi(z)}{\mu(z)}\right)_{|t=0} = \frac{\partial}{\partial t}\nu_t^\pi(F)_{|t=0} = \sum_{(x,y)\in V^2}\pi(x,y)\frac{\partial}{\partial t}\nu_t^{x,y}(F)\\ &= \sum_{(x,y)\in V^2}\pi(x,y)d(x,y)\nabla_{x,y}F(x).\end{aligned}$$

By the definition of the gradient, for any $\gamma \in \Gamma(x,y)$, it holds $\nabla_\gamma F(x) = \nabla_\gamma^+ F(x)$. Thus, by the definition of $\nabla_{x,y}F$, we get

$$\frac{\partial}{\partial t}H(\nu_t^\pi|\mu)_{|t=0} = \sum_{(x,y)\in V^2}\frac{\pi(x,y)d(x,y)}{|\Gamma(x,y)|}\sum_{\gamma\in\Gamma(x,y)}\nabla_\gamma^+F(x).$$

Now, observe that for $(x,y)\in V^2$ given, it holds

$$\sum_{\gamma\in\Gamma(x,y)}\nabla_\gamma^+F(x) = \sum_{\gamma\in\Gamma(x,y)}F(\gamma^+(x)) - F(x) = \sum_{z\sim x}(F(z)-F(x))|\Gamma(x,z,y)|\,,$$

completing the proof. □

2.4. **Tensoring property.** In this section we prove that the path $(\nu_t^{x,y})_{t\in[0,1]}$ constructed in Section 2.1 does tensorise. This will appear to be crucial in deriving the displacement convexity of the entropy on product spaces. Moreover we shall prove that, in order to have this tensoring property, the law of the random variable N_t introduced in the construction of the path $(\nu_t^{x,y})_{t\in[0,1]}$, must be, modulo a change of time, a binomial (see Proposition 2.12 below). The tensoring property of the path $(\nu_t^{x,y})_{t\in[0,1]}$ is the following.

Lemma 2.10. *Let $G_1 = (V_1,E_1)$, $G_2 = (V_2,E_2)$ be two graphs and let $G = G_1\square G_2$ be their Cartesian product. Then, for any $x = (x_1,x_2)$, $y = (y_1,y_2)$ and $z = (z_1,z_2)$ in $V_1\times V_2$,*

$$\nu_t^{x,y}(z) = \nu_t^{x_1,y_1}(z_1)\nu_t^{x_2,y_2}(z_2).$$

Proof. Fix $x = (x_1,x_2)$, $y = (y_1,y_2)$ and $z = (z_1,z_2)$ in $V_1\times V_2$. Then, we observe that, given two geodesics, one from x_1 to y_1, and one from x_2 to y_2, one can construct exactly $C_{d(x,y)}^{d(x_1,y_1)}$ different geodesics from x to y (by choosing the $d(x_1,y_1)$ positions where to change the first coordinate, according to the geodesic joining x_1 to y_1, and thus changing the second coordinate in the remaining $d(x_2,y_2) = d(x,y) - d(x_1,y_1)$ positions, according to the geodesic joining x_2 to y_2). This construction exhausts all the geodesics from x to y. Hence,

$$|\Gamma(x,y)| = C_{d(x,y)}^{d(x_1,y_1)}|\Gamma(x_1,y_1)|\times|\Gamma(x_2,y_2)|. \tag{2.11}$$

Observe also that z belongs to some geodesic from x to y if and only if z_1 and z_2 belong respectively to some geodesic from x_1 to y_1, and from x_2 to y_2. Therefore, by (2.1), it follows that

$$|\Gamma(x,z,y)| = C_{d(x,z)}^{d(x_1,z_1)}C_{d(z,y)}^{d(z_1,y_1)}|\Gamma(x_1,z_1,y_1)|\times|\Gamma(x_2,z_2,y_2)|.$$

So, it holds that

$$\begin{aligned}\nu_t^{x,y}(z) &= C_{d(x,y)}^{d(x,z)} t^{d(x,z)}(1-t)^{d(y,z)}\,\frac{|\Gamma(x,z,y)|}{|\Gamma(x,y)|}\\ &= \frac{C_{d(x,y)}^{d(x,z)} C_{d(x,z)}^{d(x_1,z_1)} C_{d(y,z)}^{d(y_1,z_1)}}{C_{d(x,y)}^{d(x_1,y_1)}} t^{d(x_1,z_1)}(1-t)^{d(y_1,z_1)} \frac{|\Gamma(x_1,z_1,y_1)|}{|\Gamma(x_1,y_1)|} t^{d(x_2,z_2)}(1-t)^{d(y_2,z_2)} \frac{|\Gamma(x_2,z_2,y_2)|}{|\Gamma(x_2,y_2)|}\\ &= \nu_t^{x_1,y_1}(z_1)\nu_t^{x_2,y_2}(z_2)\,,\end{aligned}$$

where we used that $d(x,z) = d(x_1,z_1) + d(x_2,z_2)$, and similarly for $d(y,z)$, and the fact (that the reader can easily verify) that

$$\frac{C_{d(x,y)}^{d(x,z)} C_{d(x,z)}^{d(x_1,z_1)} C_{d(y,z)}^{d(y_1,z_1)}}{C_{d(x,y)}^{d(x_1,y_1)}} = C_{d(x_1,y_1)}^{d(x_1,z_1)} C_{d(x_2,y_2)}^{d(x_2,z_2)}.$$

□

Proposition 2.12. *In the construction of* $\nu_t^{x,y}$, $t \in [0,1]$, *use a general random variable* $N_t^{d(x,y)} \in \{0,1,\ldots,d(x,y)\}$, *of parameter* $d(x,y)$ *and* t, *that satisfies a.s.* $N_0^{d(x,y)} = 0$ *and* $N_1^{d(x,y)} = d(x,y)$ *(instead of the Binomial, observe that this condition is here to ensure that* $\nu_0^{x,y} = \delta_x$ *and* $\nu_1^{x,y} = \delta_y$, *namely that* $\nu_t^{x,y}$ *is still an interpolation between the two Dirac measures) , so that*

$$\nu_t^{x,y}(z) = \mathbb{P}\left(N_t^{d(x,y)} = d(x,z)\right)\frac{|\Gamma(x,z,y)|}{|\Gamma(x,y)|}.$$

Let $G_1 = (V_1,E_1)$, $G_2 = (V_2,E_2)$ *be two graphs and let* $G = G_1 \square G_2$ *be their Cartesian product. Assume that for any* $x = (x_1,x_2)$, $y = (y_1,y_2)$ *and* $z = (z_1,z_2)$ *in* $V_1 \times V_2$,

$$\nu_t^{x,y}(z) = \nu_t^{x_1,y_1}(z_1)\nu_t^{x_2,y_2}(z_2) \qquad \forall t \in [0,1].$$

Then, there exists a function $a\colon [0,1] \to [0,1]$ *with* $a(0) = 0$, $a(1) = 1$, *such that* $N_t^{d(x,y)} \sim \mathcal{B}(a(t), d(x,y))$.

Proof. Following the proof of Lemma 2.10 we have,

$$\begin{aligned}\nu_t^{x,y}(z) &= \mathbb{P}\left(N_t^{d(x,y)} = d(x,z)\right)\frac{|\Gamma(x,z,y)|}{|\Gamma(x,y)|}\\ &= \frac{C_{d(x,z)}^{d(x_1,z_1)} C_{d(y,z)}^{d(y_1,z_1)}}{C_{d(x,y)}^{d(x_1,y_1)}}\mathbb{P}\left(N_t^{d(x,y)} = d(x,z)\right)\frac{|\Gamma(x_1,z_1,y_1)|}{|\Gamma(x_1,y_1)|}\,\frac{|\Gamma(x_2,z_2,y_2)|}{|\Gamma(x_2,y_2)|}\,.\end{aligned}$$

On the other hand,

$$\nu_t^{x_1,y_1}(z_1) = \mathbb{P}\left(N_t^{d(x_1,y_1)} = d(x_1,z_1)\right)\frac{|\Gamma(x_1,z_1,y_1)|}{|\Gamma(x_1,y_1)|}$$

and

$$\nu_t^{x_2,y_2}(z_2) = \mathbb{P}\left(N_t^{d(x_2,y_2)} = d(x_2,z_2)\right)\frac{|\Gamma(x_2,z_2,y_2)|}{|\Gamma(x_2,y_2)|}.$$

Hence, the identity $\nu_t^{x,y}(z) = \nu_t^{x_1,y_1}(z_1)\nu_t^{x_2,y_2}(z_2)$ ensures that

$$\frac{C_{d(x,z)}^{d(x_1,z_1)} C_{d(y,z)}^{d(y_1,z_1)}}{C_{d(x,y)}^{d(x_1,y_1)}}\mathbb{P}\left(N_t^{d(x,y)} = d(x,z)\right) = \mathbb{P}\left(N_t^{d(x_1,y_1)} = d(x_1,z_1)\right)\mathbb{P}\left(N_t^{d(x_2,y_2)} = d(x_2,z_2)\right)$$

for any $z_1 \in [\![x_1,y_1]\!]$, $z_2 \in [\![x_2,y_2]\!]$.

Now, observe that

$$\frac{C_{d(x,z)}^{d(x_1,z_1)} C_{d(y,z)}^{d(y_1,z_1)}}{C_{d(x,y)}^{d(x_1,y_1)}} = \frac{C_{d(x_1,y_1)}^{d(x_1,z_1)} C_{d(x_2,y_2)}^{d(x_2,z_2)}}{C_{d(x,y)}^{d(x,z)}}.$$

Hence, the latter can be rewritten as

$$\frac{\mathbb{P}\left(N_t^{d(x,y)} = d(x,z)\right)}{C_{d(x,y)}^{d(x,z)}} = \frac{\mathbb{P}\left(N_t^{d(x_1,y_1)} = d(x_1,z_1)\right)}{C_{d(x_1,y_1)}^{d(x_1,z_1)}} \times \frac{\mathbb{P}\left(N_t^{d(x_2,y_2)} = d(x_2,z_2)\right)}{C_{d(x_2,y_2)}^{d(x_2,z_2)}}.$$

Set, for simplicity, for any n, k, $0 \leq k \leq n$

$$p_{n,k} := \frac{\mathbb{P}\left(N_t^n = k\right)}{C_n^k}.$$

Notice that $p_{n,k}$ depends also on t, while not explicitly stated. We end up with the following induction formula

$$p_{n,k} = p_{n_1,k_1} \cdot p_{n-n_1,k-k_1} \tag{2.13}$$

for any integers k_1, n_1, k, n satisfying the following conditions

$$k, n_1 \leq n, \qquad k_1 \leq \min(k, n_1), \qquad \text{and} \quad n_1 - k_1 \leq n - k.$$

(We set, $n = d(x,y)$, $n_1 = d(x_1,y_1)$, $k = d(x,z)$ and $k_1 = d(x_1,z_1)$).

The special choice $n_1 = 1$, $k_1 = 0$ leads to

$$p_{n,k} = p_{1,0} \cdot p_{n-1,k}. \tag{2.14}$$

Hence, it cannot be that $p_{1,0} = 0$ (otherwise we would have $p_{n,k} = 0$ for any $k \geq 0$, any $n \geq 1$, which clearly is impossible since $\sum_{k=0}^n C_n^k p_{n,k} = 1$).

Set $b = b(t) = p_{1,0}$. From (2.14) we deduce that

$$p_{n,k} = b^{n-k} p_{k,k}.$$

Finally, the special choice $n = k$, $n_1 = k_1 = k - 1$, in (2.13), ensures that

$$p_{k,k} = p_{k-1,k-1} \cdot p_{1,1}.$$

Since $p_{1,0} + p_{1,1} = 1$, the latter reads as

$$p_{k,k} = p_{1,1}^k = (1-b)^k.$$

It follows that

$$p_{n,k} = b^{n-k}(1-b)^k \qquad \forall n,\ \forall k \leq n.$$

Now set $a(t) = 1 - b(t)$ to end up with

$$\mathbb{P}\left(N_t^n = k\right) = C_n^k a^k (1-a)^{n-k},$$

which guarantees that $N_t^{d(x,y)}$ is indeed a binomial variable of parameter $a(t)$ and $d(x,y)$.

To end the proof, it is suffices to observe that $N_0^{d(x,y)} = 0$ implies $a(0) = 0$, and that $N_1^{d(x,y)} = d(x,y)$ implies $a(1) = 1$. □

2.5. **Examples.** In this section we collect some elementary facts on specific examples. Namely we give explicit expressions of $\nu_t^{x,y}$, and derive some properties, when available, on the complete graph, the two-point space, and the hypercube.

2.5.1. *Complete graph K_n.* Let K_n be the complete graph with n vertices. Then, given any two points $x, y \in K_n$, there exists only one geodesic from x to y, namely $\Gamma(x, y) = \{(x, y)\}$. Hence, by construction of $\nu_t^{x,y}$, we have

$$\nu_t^{x,y}(z) = 0 \ \forall z \neq x, y; \quad \nu_t^{x,y}(x) = 1 - t, \quad \text{and} \quad \nu_t^{x,y}(y) = t. \tag{2.15}$$

Therefore, for any coupling π with marginals ν_0 and ν_1 (two given probability measures on K_n), we have for any $z \in K_n$,

$$\begin{aligned}\nu_t^\pi(z) &= \sum_{(x,y)\in C(z)} \nu_t^{x,y}(z)\pi(x, y) = \sum_{y\in K_n} \nu_t^{z,y}(z)\pi(z, y) + \sum_{x\in K_n} \nu_t^{x,z}(z)\pi(x, z) \\ &= (1 - t) \sum_{y\in K_n} \pi(z, y) + t \sum_{x\in K_n} \pi(x, z) = (1 - t)\nu_0(z) + t\nu_1(z).\end{aligned}$$

As a conclusion, on the complete graph, ν_t^π is a simple linear combination of ν_0 and ν_1 that does not depend on π.

Moreover, under the assumption of Corollary 2.8, since $d(x, y) = |\Gamma(x, y)| = |\Gamma(z, y)| = 1$, we have

$$\frac{\partial}{\partial t} H(\nu_t^\pi|\mu)_{|_{t=0}} = \sum_{x\in K_n}\sum_{z\sim x}(\log f(z) - \log f(x))\pi(x, z) = \sum_{z\in K_n} \log f(z)\nu_1(z) - \sum_{x\in K_n} f(x) \log f(x)\mu(x)$$

where we set for simplicity $f = \nu_0/\mu$. On the other hand, since f is a density with respect to μ,

$$\begin{aligned}-\mathcal{E}_\mu(f, \log f) &:= -\frac{1}{2} \sum_{x,z\in K_n} (\log f(z) - \log f(x))(f(z) - f(x))\mu(x)\mu(z) \\ &= \sum_{z\in K_n} \log f(z)\mu(z) - \sum_{x\in K_n} f(x) \log f(x)\mu(x).\end{aligned}$$

Hence, if $\nu_1 = \mu \equiv 1/n$ is the uniform measure on K_n (notice all the measures on K_n are then absolutely continuous with respect to μ), we can conclude that

$$\frac{\partial}{\partial t} H(\nu_t^\pi|\mu)_{|_{t=0}} = -\mathcal{E}_\mu(f, \log f). \tag{2.16}$$

Note that, when $\mu \equiv 1/n$, $\mathcal{E}_\mu$ corresponds to the Dirichlet form associated to the uniform chain on the complete graph (each point can jumps to each point with probability $1/n$).

As a summary, on the complete graph we have: For any coupling π, for any $t \in [0, 1]$,

$$\nu_t^\pi = (1 - t)\nu_0 + t\nu_1.$$

For $\nu_1 = \mu \equiv 1/n$ and $f = \nu_0/\mu$, it holds

$$\frac{\partial}{\partial t} H(\nu_t^\pi|\mu)_{|_{t=0}} = -\mathcal{E}_\mu(f, \log f).$$

2.5.2. *The two-point space.* The previous computations apply in particular to the two-point space $\{0, 1\}$. In this specific case, let us consider μ to be a Bernoulli(p) measure (*i.e.* $\mu(1) = p = 1 - q = 1 - \mu(0)$). As above, $\nu_t^\pi = (1 - t)\nu_0 + t\nu_1$, for any coupling π of ν_0 and ν_1. Moreover, it can also be checked by an easy computation that, for any $t \in [0, 1]$,

$$\frac{\partial^2}{\partial t^2} H(\nu_t^\pi|\mu) = \frac{C^2}{(\nu_0(0) + tC)(\nu_0(1) - tC)} \geq 4C^2 ,$$

where $C = \nu_1(0) - \nu_0(0)$, and $\|\nu_0 - \nu_1\|_{TV} = |\nu_1(0) - \nu_0(0)|$. As a result, one arrives at the following displacement convexity of the entropy of ν_t^π on the two-point space:

$$H(\nu_t^\pi|\mu) \le (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - 2t(1-t)\|\nu_0 - \nu_1\|_{TV}^2, \qquad t \in [0,1]\,. \tag{2.17}$$

In Section 4 below, we refine the above inequality further, and generalize in two ways – by deriving displacement convexity of entropy on the complete graph and the n-dimensional hypercube.

As an application, let us set $\nu_1 = \mu$, and use $f = \nu_0/\mu$ for the density; taking the limit $t \to 0$, and using

$$\frac{\partial}{\partial t} H(\nu_t^\pi|\mu)_{|_{t=0}} = -\frac{pq}{2}(f(1) - f(0))(\log f(1) - \log f(0)) =: -\mathcal{E}_\mu(f, \log f)\,,$$

we get a *reinforced* modified logarithmic Sobolev inequality on the two-point space of the following type:

$$\mathrm{Ent}_\mu(f) \le \mathcal{E}_\mu(f, \log f) - 2\|f\mu - \mu\|_{TV}^2\,. \tag{2.18}$$

In the above, $\mathcal{E}_\mu(f, \log f)$ corresponds to the Dirichlet form associated with the Markov chain jumping from 0 to 1 with probability p and from 1 to 0 with probability q. The inequality is a reinforcement of a modified log-Sobolev inequality, considered by previous researchers (as mentioned in the introduction), which lacks the negative term. Similarly to (2.17), we also refine (2.18) further in Proposition 5.12.

2.5.3. *The n-dimensional hypercube* Ω_n. Consider the n-dimensional hypercube $\Omega_n = \{0,1\}^n$ whose edges consist of pairs of vertices p that differ in precisely one coordinate. The graph distance here coincides with the Hamming distance:

$$d(x,y) = \sum_{i=1}^n \mathbb{1}_{x_i \neq y_i}, \qquad x, y \in \Omega_n.$$

Then, one observes that $|\Gamma(x,y)| = d(x,y)!$ (since, in order to move from x to y in the shortest way, one just needs to choose, among $d(x,y)$ coordinates where x and y differ, the order of the flips (*i.e.* moves from x_i to $1 - x_i$)). It follows from (2.2) that, as soon as z belongs to a geodesic from x to y,

$$\nu_t^{x,y}(z) = C_{d(x,y)}^{d(x,z)} t^{d(x,z)} (1-t)^{d(y,z)} \frac{d(x,z)!d(y,z)!}{d(x,y)!} = t^{d(x,z)}(1-t)^{d(y,z)},$$

and $\nu_t^{x,y}(z) = 0$ if z does not belong to a geodesic from x to y.

This expression can be recovered using the tensorisation property above. Namely, observe that Equation (2.15) can be rewritten for the two-point space as follows, for all coordinates:

$$\nu_t^{x_i,y_i}(z_i) = \mathbb{1}_{\{x_i,y_i\}}(z_i) t^{d(x_i,z_i)}(1-t)^{d(y_i,z_i)}.$$

Hence, by Lemma 2.10,

$$\nu_t^{x,y}(z) = \prod_{i=1}^n \nu_t^{x_i,y_i}(z_i) = t^{d(x,z)}(1-t)^{d(y,z)}\,,$$

as soon as z belongs to a geodesic from x to y, and 0 otherwise. Observe that the latter can also be rewritten in terms of a product of probability measures on the fibers as

$$\nu_t^{x,y} = \otimes_{i=1}^n ((1-t)\delta_{x_i} + t\delta_{y_i}). \tag{2.19}$$

Given two probability measures on Ω_n, and a coupling π on $\Omega_n \times \Omega_n$, we can finally define

$$\nu_t^\pi(z) = \sum_{(x,y)\in\Omega_n^2} t^{d(x,z)}(1-t)^{d(y,z)}\pi(x,y).$$

On the n-dimensional hypercube we have: for any couple $(x,y) \in \Omega_n^2$ and for any $t \in [0,1]$,

$$\nu_t^{x,y} = \sum_{z \in [\![x,y]\!]} t^{d(x,z)}(1-t)^{d(y,z)}\delta_z = \otimes_{i=1}^n ((1-t)\delta_{x_i} + t\delta_{y_i}).$$

3. Weak transport cost

In this section we recall a notion of a discrete Wasserstein-type distance, called weak transport cost – introduced and studied in [33, 52], developed further in [17] – and collect some useful facts from [17]. Also, we introduce the notion of a Knothe-Rosenblatt coupling which will play a crucial role in the displacement convexity of the entropy property on product spaces.

3.1. Definition and first properties.

For the notion of a weak transport cost, first recall the definition of $P(\nu_0,\nu_1)$ introduced in Section 1.1.

Definition 3.1. *Let $\nu_0,\nu_1 \in \mathcal{P}(V)$. Then, the weak transport cost $\widetilde{\mathcal{T}}_2(\nu_1|\nu_0)$ between ν_0 and ν_1 is defined as*

$$\widetilde{\mathcal{T}}_2(\nu_1|\nu_0) := \inf_{p \in P(\nu_0,\nu_1)} \sum_{x\in V}\left(\sum_{y\in V} d(x,y)p(x,y)\right)^2 \nu_0(x).$$

It can be shown that

$$(\nu_0,\nu_1) \mapsto \sqrt{\widetilde{\mathcal{T}}_2(\nu_1|\nu_0)} + \sqrt{\widetilde{\mathcal{T}}_2(\nu_0|\nu_1)}$$

is a distance on $\mathcal{P}(V)$, see [17].

Also recall from the introduction, the following notation: given $\pi \in \Pi(\nu_0,\nu_1)$, consider the kernels $p \in P(\nu_0,\nu_1)$ and $\bar{p} \in P(\nu_1,\nu_0)$ defined by $\pi(x,y) = \nu_0(x)p(x,y) = \nu_1(y)\bar{p}(y,x)$ and set

$$I_2(\pi) := \sum_{x\in V}\left(\sum_{y\in V} d(x,y)p(x,y)\right)^2 \nu_0(x), \tag{3.2}$$

$$\bar{I}_2(\pi) := \sum_{y\in V}\left(\sum_{x\in V} d(x,y)\bar{p}(y,x)\right)^2 \nu_1(y),$$

and

$$J_2(\pi) := \left(\sum_{x\in V}\sum_{y\in V} d(x,y)\pi(x,y)\right)^2.$$

With this notation,

$$\widetilde{\mathcal{T}}_2(\nu_0|\nu_1) = \inf_{\pi\in\Pi(\nu_0,\nu_1)} I_2(\pi).$$

Also, define

$$\hat{\mathcal{T}}_2(\nu_0,\nu_1) := \inf_{\pi\in\Pi(\nu_0,\nu_1)} J_2(\pi),$$

and observe that $\hat{\mathcal{T}}_2(\nu_0,\nu_1) = W_1^2(\nu_0,\nu_1)$ where W_1 is the usual L_1-Wasserstein distance associated to the distance d.

When ν_0 and ν_1 are absolutely continuous with respect to some probability measure μ, and d is the Hamming distance $d(x,y) = \mathbb{1}_{x\neq y}$, $x,y \in V$, the weak transport cost and the L_1-Wasserstein distance take an explicit form. This is stated in the next lemma. We give the proof for completeness.

Lemma 3.3 ([17])**.** *Assume that $\nu_0, \nu_1 \in \mathcal{P}(V)$ are absolutely continuous with respect to a third probability measure $\mu \in \mathcal{P}(V)$, with respective densities f_0 and f_1. Assume that $d(x,y) = \mathbb{1}_{x \neq y}$, $x, y \in V$. Then it holds*

$$\widetilde{\mathcal{T}}_2(\nu_1|\nu_0) = \int \left[1 - \frac{f_1}{f_0}\right]_+^2 f_0 \, d\mu$$

where $[X]_+ = \max(X, 0)$, and

$$\sqrt{\hat{\mathcal{T}}_2(\nu_0, \nu_1)} = \int [f_0 - f_1]_+ \, d\mu = \frac{1}{2} \int |f_0 - f_1| \, d\mu = \frac{1}{2} \|\nu_0 - \nu_1\|_{TV}$$

with $\| \cdot \|_{TV}$, the total variation norm.

Remark 3.4. *Observe that $\widetilde{\mathcal{T}}_2(\nu_1|\nu_0)$ does not depend on μ.*

Proof. For any $\pi \in \Pi(\nu_0, \nu_1)$ and any $x \in V$, one has

$$1 - \sum_{y \in V} d(x,y) p(x,y) = \frac{\pi(x,x)}{\nu_0(x)} \leq \frac{\min(\nu_0(x), \nu_1(x))}{\nu_0(x)} = \min\left(\frac{f_1(x)}{f_0(x)}, 1\right).$$

and therefore

$$\left[1 - \frac{f_1(x)}{f_0(x)}\right]_+ \leq \sum_{y \in V} d(x,y) p(x,y).$$

By integrating with respect to the measure ν_0 and then optimizing over all $\pi \in \Pi(\nu_0, \nu_1)$, it follows that

$$\int [f_0 - f_1]_+ \, d\mu \leq \sqrt{\hat{\mathcal{T}}_2(\nu_0, \nu_1)},$$

and

$$\int \left[1 - \frac{f_1}{f_0}\right]_+^2 f_0 \, d\mu \leq \widetilde{\mathcal{T}}_2(\nu_1|\nu_0).$$

The equality is reached choosing $\pi^* \in \Pi(\nu_0, \nu_1)$ defined by

$$\pi^*(x,y) = \nu_0(x) p^*(x,y) = \mathbb{1}_{x=y} \min(\nu_0(x), \nu_1(x)) + \mathbb{1}_{x \neq y} \frac{[\nu_0(x) - \nu_1(x)]_+ [\nu_1(y) - \nu_0(y)]_+}{\sum_{z \in V} [\nu_1(z) - \nu_0(z)]_+}, \tag{3.5}$$

since $\sum_{y \in V} d(x,y) p^*(x,y) = \left[1 - \frac{f_1(x)}{f_0(x)}\right]_+$. □

3.2. **The Knothe-Rosenblatt coupling.** In this subsection, we recall a general method, due to Knothe-Rosenblatt [24, 47], enabling to construct couplings between probability measures on product spaces.

Consider two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ and two probability measures $\nu_0, \nu_1 \in \mathcal{P}(V_1 \times V_2)$. The disintegration formulas of ν_0, ν_1 (recall (1.8)) read

$$\nu_0(x_1, x_2) = \nu_0^2(x_2) \nu_0^1(x_1|x_2) \qquad \text{and} \qquad \nu_1(y_1, y_2) = \nu_1^2(y_2) \nu_1^1(y_1|y_2). \tag{3.6}$$

Let $\pi^2 \in \mathcal{P}(V_2^2)$ be a coupling of ν_0^2, ν_1^2, and for all $(x_2, y_2) \in V_2^2$ let $\pi^1(\cdot|x_2, y_2) \in \mathcal{P}(V_1^2)$ be a coupling of $\nu_0^1(\cdot|x_2)$ and $\nu_1^1(\cdot|y_2)$, $x_2, y_2 \in V_2$. We are now in a position to define the Knothe-Rosenblatt coupling.

Definition 3.7 (Knothe-Rosenblatt coupling)**.** *Let* $\nu_0, \nu_1 \in \mathcal{P}(V_1 \times V_2)$*, and consider a family of couplings* $\pi^2, \{\pi^1(\,\cdot\,|x_2, y_2)\}_{x_2,y_2}$ *as above; the coupling* $\hat{\pi} \in \mathcal{P}([V_1 \times V_2]^2)$*, defined by*

$$\hat{\pi}((x_1, x_2), (y_1, y_2)) := \pi^2(x_2, y_2)\pi^1(x_1, y_1|x_2, y_2)\,, \qquad (x_1, x_2), (y_1, y_2) \in V_1 \times V_2$$

is called the Knothe-Rosenblatt coupling *of* ν_0, ν_1 *associated with the family of couplings*

$$\left\{\pi^2, \{\pi^1(\,\cdot\,|x_2, y_2)\}_{x_2,y_2}\right\}.$$

It is easy to check that the Knothe-Rosenblatt coupling is indeed a coupling of ν_0, ν_1. Note that it is usually required that the couplings $\pi^2, \{\pi^1(\,\cdot\,|x_2, y_2)\}_{x_2,y_2}$ are optimal for some weak transport cost, but we will not make this assumption in what follows.

The preceding construction can easily be generalized to products of n graphs. Consider n graphs $G_1 = (V_1, E_1), \ldots, G_n = (V_n, E_n)$, and two probability measures $\nu_0, \nu_1 \in \mathcal{P}(V_1 \times \cdots \times V_n)$ admitting the following disintegration formulas: for all $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in V_1 \times \cdots \times V_n$,

$$\begin{aligned} \nu_0(x) &= \nu_0^n(x_n)\nu_0^{n-1}(x_{n-1}|x_n)\nu_0^{n-2}(x_{n-2}|x_{n-1}, x_n) \cdots \nu_0^1(x_1|x_2, \ldots, x_n), \\ \nu_1(y) &= \nu_1^n(y_n)\nu_1^{n-1}(y_{n-1}|y_n)\nu_1^{n-2}(y_{n-2}|y_{n-1}, y_n) \cdots \nu_1^1(y_1|y_2, \ldots, y_n). \end{aligned}$$

For all $j = 1, \ldots, n$, let $\pi^j(\,\cdot\,|x_{j+1}, \ldots, x_n, y_{j+1}, \ldots, y_n) \in \mathcal{P}(V_j^2)$ be a coupling of $\nu_0^j(\,\cdot\,|x_{j+1}, \ldots, x_n)$ and $\nu_1^j(\,\cdot\,|y_{j+1}, \ldots, y_n)$. The Knothe-Rosenblatt coupling $\hat{\pi} \in \mathcal{P}([V_1 \times \cdots \times V_n]^2)$ between ν_0 and ν_1 is then defined by

$$\hat{\pi}(x, y) = \pi^n(x_n, y_n)\pi^{n-1}(x_{n-1}, y_{n-1}|x_n, y_n) \cdots \pi^1(x_1, y_1|x_2, \ldots, x_n, y_2, \ldots, y_n),$$

for all $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$.

3.3. **Tensorisation.** Another useful property of the weak transport cost defined above is that it tensorises in the following sense. For $1 \le i \le n$, let $G_i = (V_i, E_i)$ be a graph with the associated distance d_i. Given two probability measures ν_0, ν_1 in $\mathcal{P}(V_1 \times \cdots \times V_n)$, define

$$\widetilde{\mathcal{T}}_2^{(n)}(\nu_1|\nu_0) := \inf_{p \in P(\nu_0,\nu_1)} \sum_{x \in V_1 \times \cdots \times V_n} \sum_{i=1}^n \left(\sum_{y \in V_1 \times \cdots \times V_n} d_i(x_i, y_i)p(x, y) \right)^2 \nu_0(x)$$

where $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in V_1 \times \cdots \times V_n$.

As above, for any coupling π of $\nu_0, \nu_1 \in \mathcal{P}(V_1 \times \cdots \times V_n)$ we also define

$$I_2^{(n)}(\pi) := \sum_{x \in V_1 \times \cdots \times V_n} \sum_{i=1}^n \left(\sum_{y \in V_1 \times \cdots \times V_n} d_i(x_i, y_i)p(x, y) \right)^2 \nu_0(x)$$

where p is such that $\pi(x, y) = \nu_0(x)p(x, y)$, for all $x, y \in V_1 \times \cdots \times V_n$. Similarly, one defines $\bar{I}_2^{(n)}$.

We also define

$$J_2^{(n)}(\pi) := \sum_{i=1}^n \left(\sum_{x,y \in V_1 \times \cdots \times V_n} d_i(x_i, y_i)\pi(x, y) \right)^2$$

and

$$\hat{\mathcal{T}}_2^{(n)}(\nu_0, \nu_1) := \inf_{\pi \in \Pi(\nu_0,\nu_1)} J_2^{(n)}(\pi).$$

Using the notation of Section 3.2 above, we can state the result.

Proposition 3.8. *Let ν_0, ν_1 in $\mathcal{P}(V_1 \times \cdots \times V_n)$; and consider a family of couplings $\pi^n \in \Pi(\nu_0^n, \nu_1^n)$ and $\pi^k(\,\cdot\,|x_{k+1}, \dots, x_n) \in \Pi(\nu_0^k(\,\cdot\,|x_{k+1}, \dots, x_n), \nu_1^k(\,\cdot\,|y_{k+1}, \dots, y_n))$ with $(x_2, \dots, x_n), (y_2, \dots, y_n) \in V_2 \times \cdots \times V_n$, as above. Then,*

$$I_2^{(n)}(\hat{\pi}) \leq I_2(\pi^n) + \sum_{k=1}^{n-1} \sum_{x,y \in V_1 \times \cdots \times V_n} \hat{\pi}(x,y) I_2(\pi^k(\,\cdot\,|x_{k+1}, \dots, x_n, y_{k+1} \dots y_n)).$$

where $\hat{\pi}$ is the Knothe-Rosenblatt coupling of ν_0 and ν_1 associated with the family of couplings above. The same holds for $\bar{I}_2^{(n)}$ and $J_2^{(n)}(\pi)$.

In particular, if the couplings π^n and $\pi^k(\,\cdot\,|x_{k+1}, \dots, x_n)$ are assumed to achieve the infimum in the definition of the weak transport costs between ν_0^n and ν_1^n and between $\nu_0^k(\,\cdot\,|x_{k+1}, \dots, x_n)$ and $\nu_1^k(\,\cdot\,|y_{k+1}, \dots, y_n)$ for all $k \in \{1, \dots, n-1\}$, we immediately get the following tensorisation inequality for $\widetilde{\mathcal{T}}_2$:

$$\widetilde{\mathcal{T}}_2^{(n)}(\nu_1|\nu_0) \leq \widetilde{\mathcal{T}}_2(\nu_1^n|\nu_0^n) + \sum_{k=1}^{n-1} \sum_{\substack{x,y \in \\ V_1 \times \cdots \times V_n}} \hat{\pi}(x,y) \widetilde{\mathcal{T}}_2(\nu_1^k(\cdot|x_{k+1}, \dots, x_n)|\nu_0^k(\cdot|y_{k+1}, \dots, y_n)). \tag{3.9}$$

In an obvious way, the same kind of conclusion holds replacing $\widetilde{\mathcal{T}}_2$ by $\hat{\mathcal{T}}_2$.

Proof. In this proof, we will use the following shorthand notation: if $x \in V$ and if $1 \leq i \leq j \leq n$, we will denote by $x_{i:j}$ the subvector $(x_i, x_{i+1}, \dots, x_j) \in V_i \times \cdots \times V_j$.

Define the kernels $\hat{p}(\,\cdot\,,\,\cdot\,)$, $p^n(\,\cdot\,,\,\cdot\,)$ and $p^k(\,\cdot\,,\,\cdot\,|x_{k+1:n}, y_{k+1:n})$ by the formulas

$$\begin{aligned}
\hat{\pi}(x,y) &= \hat{p}(x,y)\nu_0(x) \\
\pi^k(x_k, y_k|x_{k+1:n}, y_{k+1:n}) &= p^k(x_k, y_k|x_{k+1:n}, y_{k+1:n})\nu_0^k(x_k|x_{k+1:n}), \quad \forall k < n, \\
\pi^n(x_n, y_n) &= p^n(x_n, y_n)\nu_0^n(x_n).
\end{aligned}$$

By the definition of the Knothe-Rosenblatt coupling $\hat{\pi}$, it holds

$$\hat{p}(x,y) = \prod_{k=1}^{n-1} p^k(x_k, y_k|x_{k+1:n}, y_{k+1:n}) \times p^n(x_n, y_n).$$

As a result,

$$\begin{aligned}
\left(\sum_y d_i(x_i, y_i)\hat{p}(x,y) \right)^2 &= \left(\sum_{y_{i:n}} d_i(x_i, y_i) \prod_{k=i}^{n-1} p^k(x_k, y_k|x_{k+1:n}, y_{k+1:n}) p^n(x_n, y_n) \right)^2 \\
&\leq \sum_{y_{i+1:n}} \prod_{k=i+1}^{n-1} p^k(x_k, y_k|x_{k+1:n}, y_{k+1:n}) p^n(x_n, y_n) \left(\sum_{y_i} d_i(x_i, y_i) p^i(x_i, y_i|x_{i+1:n}, y_{i+1:n}) \right)^2
\end{aligned}$$

where the inequality comes from Jensen's inequality. Therefore,

$$\begin{aligned}
&\sum_x \left(\sum_y d_i(x_i,y_i)\hat{p}(x,y)\right)^2 \nu_0(x)\\
&\le \sum_{x_{i+1:n}}\sum_{y_{i+1:n}}\prod_{k=i+1}^{n-1}\pi^k(x_k,y_k|x_{k+1:n},y_{k+1:n})\pi^n(x_n,y_n)\sum_{x_i}\nu_0^i(x_i|x_{i+1:n})\left(\sum_{y_i} d_i(x_i,y_i)p^i(x_i,y_i|x_{i+1:n},y_{i+1:n})\right)^2\\
&= \sum_{x_{i+1:n}}\sum_{y_{i+1:n}}\prod_{k=i+1}^{n-1}\pi^k(x_k,y_k|x_{k+1:n},y_{k+1:n})\pi^n(x_n,y_n)I_2(\pi^i(\,\cdot\,|x_{i+1:n},y_{i+1:n}))\\
&= \sum_{x,y}\hat{\pi}(x,y)I_2(\pi^i(\,\cdot\,|x_{i+1:n},y_{i+1:n})).
\end{aligned}$$

Similarly

$$\sum_x \left(\sum_y d_n(x_n,y_n)\hat{p}(x,y)\right)^2 \nu_0(x) \le \sum_{x,y}\hat{\pi}(x,y)I_2(\pi^n).$$

Summing all these inequalities gives the announced tensorisation formula.

The proof for $\bar{I}_2^{(n)}$ and $J_2^{(n)}$ is identical and left to the reader. □

4. Displacement convexity property of the entropy.

Using the weak transport cost defined in the previous section, we can now derive a displacement convexity property of the entropy on graphs. More precisely, we will derive such a property for the complete graph. Then we will prove that our definition of ν_t^π allows the displacement convexity to tensorise. As a consequence, we will be able to derive such a property on the n-dimensional hypercube.

4.1. The complete graph.

Consider the complete graph K_n, or equivalently any graph G equipped with the Hamming distance $d(x,y) = \mathbb{1}_{x\neq y}$ (in the definition of the weak transport cost). Recall the definition of ν_t^π given in (2.4), and that we proved, in Section 2.5.1, that $\nu_t^\pi = (1-t)\nu_0 + t\nu_1$ for any choice of coupling π. Then, the following holds.

Proposition 4.1 (Displacement convexity on the complete graph)**.** *Let ν_0 ,ν_1, $\mu \in \mathcal{P}(K_n)$ be three probability measures. Assume that ν_0,ν_1 are absolutely continuous with respect to μ. Then*

$$H(\nu_t|\mu) \le (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{t(1-t)}{2}\left(\widetilde{\mathcal{T}}_2(\nu_1|\nu_0) + \widetilde{\mathcal{T}}_2(\nu_0|\nu_1)\right), \quad \forall t\in[0,1],$$

where $\nu_t = (1-t)\nu_0 + t\nu_1$.

Proof. Our aim is simply to bound from below the second order derivative of $t \mapsto F(t) := H(\nu_t|\mu)$. Denote by f_0 and f_1 the respective densities of ν_0 and ν_1 with respect to μ. We have

$$F(t) = \int \log\left((1-t)f_0 + tf_1\right)\left((1-t)f_0 + tf_1\right)\, d\mu.$$

Thus $F'(t) = \int \log\left((1-t)f_0 + tf_1\right)\, d(\nu_0 - \nu_1)$. In turn,

$$\begin{aligned} F''(t) &= \int \frac{(f_0 - f_1)^2}{(1-t)f_0 + tf_1}\, d\mu = \int \frac{[f_0 - f_1]_+^2}{(1-t)f_0 + tf_1}\, d\mu + \int \frac{[f_1 - f_0]_+^2}{(1-t)f_0 + tf_1}\, d\mu \\ &\geq \int \frac{[f_0 - f_1]_+^2}{f_0}\, d\mu + \int \frac{[f_1 - f_0]_+^2}{f_1}\, d\mu = \int \left[1 - \frac{f_1}{f_0}\right]_+^2 f_0\, d\mu + \int \left[1 - \frac{f_0}{f_1}\right]_+^2 f_1\, d\mu \\ &= \widetilde{\mathcal{T}}_2(\nu_1|\nu_0) + \widetilde{\mathcal{T}}_2(\nu_0|\nu_1), \end{aligned}$$

where, in the last line, we used Lemma 3.3. As a consequence, the function $G\colon t \mapsto F(t) - \frac{t^2}{2}\left(\widetilde{\mathcal{T}}_2(\nu_1|\nu_0) + \widetilde{\mathcal{T}}_2(\nu_0|\nu_1)\right)$ is convex on $[0,1]$, so that $G(t) \leq (1-t)G(0) + tG(1)$ which gives precisely, after some algebra, the desired inequality. □

Remark 4.2 (Pinsker inequality)**.** *As an immediate consequence of the previous proposition, we will derive Csiszar-Kullback-Pinsker inequality (*[42, 25, 9]*). Recall the notation of the proof of Proposition 4.1. Applying Cauchy-Schwarz yields*

$$F''(t) = \int \left(\frac{|f_0 - f_1|}{\sqrt{(1-t)f_0 + tf_1}}\right)^2 d\mu \int \left(\sqrt{(1-t)f_0 + tf_1}\right)^2 d\mu \geq \left(\int |f_0 - f_1|\, d\mu\right)^2 = \|\nu_0 - \nu_1\|_{TV}^2.$$

Hence the map $G : t \mapsto F(t) - \frac{t^2}{2}\|\nu_0 - \nu_1\|_{TV}^2$ is convex on $[0,1]$ *so that*

$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{t(1-t)}{2}\|\nu_0 - \nu_1\|_{TV}^2, \qquad \forall t \in [0,1]. \tag{4.3}$$

Inequality (4.3) *is a reinforcement of the well known Csiszar-Kullback-Pinsker's inequality (see e.g.* [1, Theorem 8.2.7]*) which asserts that*

$$\|\nu_0 - \nu_1\|_{TV}^2 \leq 2H(\nu_1|\nu_0).$$

Indeed, take $\mu = \nu_0$ together with the fact that $H(\nu_t|\mu) \geq 0$, and then take the limit $t \to 0$ in (4.3) *to obtain the above inequality.*

Csiszar-Kullback-Pinsker's inequality, and its generalizations, are known to have many applications in Probability theory, Analysis and Information theory, see [57, Page 636] *for a review.*

Now we compare the displacement convexity property of Proposition 4.1 with (4.3)*. For the two-point space it is easy to check that the ratio*

$$\frac{\widetilde{\mathcal{T}}_2(\nu_1|\nu_0) + \widetilde{\mathcal{T}}_2(\nu_0|\nu_1)}{\|\nu_0 - \nu_1\|_{TV}^2}$$

is not uniformly bounded above over all probability measures ν_0 and ν_1. On the other hand, we claim that

$$\frac{\widetilde{\mathcal{T}}_2(\nu_1|\nu_0) + \widetilde{\mathcal{T}}_2(\nu_0|\nu_1)}{\|\nu_0 - \nu_1\|_{TV}^2} \geq \frac{1}{2}, \qquad \forall \nu_0, \nu_1 \tag{4.4}$$

which implies that the result in Proposition 4.1 is stronger than (4.3)*, up to a constant 2. We also provide an example below which shows that we cannot exactly recover* (4.3) *using Proposition 4.1.*

Let us prove the claim, and more precisely that the following holds

$$\widetilde{\mathcal{T}}_2(\nu_1|\nu_0) + \widetilde{\mathcal{T}}_2(\nu_0|\nu_1) \geq \frac{\|\nu_0 - \nu_1\|_{TV}^2}{1 + \frac{\|\nu_0 - \nu_1\|_{TV}}{2}} \geq \frac{1}{2}\|\nu_0 - \nu_1\|_{TV}^2. \tag{4.5}$$

This is a consequence of Cauchy-Schwarz inequality, namely, we have

$$\widetilde{\mathcal{T}}_2(\nu_1|\nu_0) + \widetilde{\mathcal{T}}_2(\nu_0|\nu_1) \geq \frac{\left(\int [f_1 - f_0]_+ d\mu\right)^2}{\nu_1(f_1 \geq f_0)} + \frac{\left(\int [f_0 - f_1]_+ d\mu\right)^2}{\nu_0(f_0 > f_1)}.$$

Since $\|\nu_0 - \nu_1\|_{TV} = 2\int [f_1 - f_0]_+ d\mu = 2(\nu_1(f_1 \geq f_0) - \nu_0(f_1 \geq f_0))$*, we get*

$$\widetilde{\mathcal{T}}_2(\nu_1|\nu_0) + \widetilde{\mathcal{T}}_2(\nu_0|\nu_1) \geq \inf_{u\in[0,1]} \frac{(1 + \frac{\|\nu_0-\nu_1\|_{TV}}{2})\|\nu_0 - \nu_1\|_{TV}^2}{4u(1 + \frac{\|\nu_0-\nu_1\|_{TV}}{2} - u)} = \frac{\|\nu_0 - \nu_1\|_{TV}^2}{1 + \frac{\|\nu_0-\nu_1\|_{TV}}{2}}.$$

We now give the example that achieves equality *in the first inequality of* (4.5)*, thus confirming that Proposition 4.1 can not exactly recover* (4.3) *: Let* ν_0 *and* ν_1 *be two probability measures on the two-point space* $\{0, 1\}$ *defined by* $\nu_1(1) = \nu_0(0) = 3/4$ *and* $\nu_1(0) = \nu_0(1) = 1/4$*. Then*

$$\|\nu_0 - \nu_1\|_{TV} = 2(\nu_1(1) - \nu_0(1)) = 1,$$

and

$$\widetilde{\mathcal{T}}_2(\nu_1|\nu_0) + \widetilde{\mathcal{T}}_2(\nu_0|\nu_1) = \frac{(\nu_1(1) - \nu_0(1))^2}{\nu_1(1)} + \frac{(\nu_0(0) - \nu_1(0))^2}{\nu_0(0)} = 2/3,$$

which gives the (claimed) equality in (4.5)*.*

4.2. **Tensorisation of the displacement convexity property.** In this section we prove that if the displacement convexity property of the entropy holds on n graphs $G_1 = (V_1, E_1), \dots, G_n = (V_n, E_n)$, equipped with probability measures $\mu_1, \dots, \mu_n$ and graph distances $d_1, \dots, d_n$ respectively, then the displacement convexity of the entropy holds on their Cartesian product equipped with $\mu_1 \otimes \cdots \otimes \mu_n$ with respect to the tensorised transport costs $I_2^{(n)}$ and $\bar{I}_2^{(n)}$. As an application we shall apply such a property to the specific example of the hypercube at the end of the section.

The next theorem is one of our main results.

Theorem 4.6. *Let* $(\mu^1, \dots, \mu^n) \in \mathcal{P}(V_1) \times \cdots \times \mathcal{P}(V_n)$*. Assume that for all* $i \in \{1, \dots, n\}$ *there is a constant* $C_i \geq 0$ *such that for all* $\nu_0, \nu_1 \in \mathcal{P}(V_i)$ *there exists* $\pi = \pi^i \in \Pi(\nu_0, \nu_1)$ *such that for all* $t \in [0, 1]$ *it holds that:*

$$H(\nu_t^\pi|\mu^i) \leq (1 - t)H(\nu_0|\mu^i) + tH(\nu_1|\mu^i) - C_i t(1 - t)(I_2(\pi) + \bar{I}_2(\pi)).$$

Then the product probability measure $\mu = \mu^1 \otimes \cdots \otimes \mu^n$ *defined on* $G = (V, E) = G_1 \,\square \cdots \square\, G_n$ *verifies the following property: for all* $\nu_0, \nu_1 \in \mathcal{P}(V)$ *there exists* $\pi = \pi^{(n)} \in \Pi(\nu_0, \nu_1)$ *such that for all* $t \in [0, 1]$ *it holds that:*

$$H(\nu_t^\pi|\mu) \leq (1 - t)H(\nu_0|\mu) + tH(\nu_1|\mu) - Ct(1 - t)(I_2^{(n)}(\pi) + \bar{I}_2^{(n)}(\pi)),$$

where $C = \min_i C_i$*. The same proposition holds replacing* $I_2(\pi) + \bar{I}_2(\pi)$ *by* $J_2(\pi)$ *and* $I_2^{(n)}(\pi) + \bar{I}_2^{(n)}(\pi)$ *by* $J_2^{(n)}(\pi)$*.*

Proof. In this proof, we use the notation and definitions introduced in Section 3.2. Fix $\nu_0, \nu_1 \in \mathcal{P}(V)$ and write the following disintegration formulas

$$\nu_0(x) = \nu_0^n(x_n) \prod_{i=1}^{n-1} \nu_0^k(x_k|x_{k+1:n}), \qquad \forall x = (x_1, \dots, x_n) \in V$$

$$\nu_1(y) = \nu_1^n(y_n) \prod_{i=1}^{n-1} \nu_1^k(y_k|y_{k+1:n}), \qquad \forall y = (y_1, \dots, y_n) \in V,$$

where we recall that $x_{k+1:n} = (x_{k+1}, \dots, x_n) \in V_{k+1} \times \cdots \times V_n$.

By assumption, for every $x, y \in V$, there are couplings $\pi^n \in \mathcal{P}(V_n \times V_n)$ and $\pi^k(\,\cdot\,|x_{k+1:n}, y_{k+1:n}) \in \mathcal{P}(V_k \times V_k)$ such that

$$\pi^n \in \Pi(\nu_0^n, \nu_1^n) \quad \text{and} \quad \pi^k(\,\cdot\,|x_{k+1:n}, y_{k+1:n}) \in \Pi(\nu_0^k(\,\cdot\,|x_{k+1:n}), \nu_1^k(\,\cdot\,|y_{k+1:n})),$$

and for which the following inequalities hold

$$\begin{aligned} H(\nu_t^n|\mu^n) &\le (1-t)H(\nu_0^n|\mu^n) + tH(\nu_1^n|\mu^n) - C_n t(1-t) J_2(\pi^n), \\ H(\nu_t^{k,x_{k+1:n},y_{k+1:n}}|\mu^k) &\le (1-t)H(\nu_0^k(\,\cdot\,|x_{k+1:n})|\mu^k) + tH(\nu_1^k(\,\cdot\,|y_{k+1:n})|\mu^k) \\ &\quad - C_k t(1-t) R_2(\pi^k(\,\cdot\,|x_{k+1:n}, y_{k+1:n})), \end{aligned}$$

where $R_2 := I_2 + \bar{I}_2$, $\nu_t^n := \nu_t^{\pi_n}$, and $\nu_t^{k,x_{k+1:n},y_{k+1:n}} = \nu_t^{\pi^k(\,\cdot\,|x_{k+1:n},y_{k+1:n})}$.

Now, consider the Knothe-Rosenblatt coupling $\hat{\pi} \in \Pi(\nu_0, \nu_1)$ constructed from the couplings π^n and $\pi^k(\,\cdot\,|x_{k+1:n}, y_{k+1:n})$, $x, y \in V$ and denote by γ_t the path $\nu_t^{\hat{\pi}} \in \mathcal{P}(V)$ connecting ν_0 to ν_1.

Let us consider the disintegration of γ_t with respect to its marginals:

$$\gamma_t(z) = \gamma_t^n(z_n)\gamma_t^{n-1}(z_{n-1}|z_n)\cdots\gamma_t^1(z_1|z_2, \dots, z_n).$$

We claim that there exist non-negative coefficients $\alpha_t^k(x_{k+1:n}, y_{k+1:n}, z_{k+1:n})$ such that

$$\sum_{x_{k+1:n}, y_{k+1:n}} \alpha_t^k(x_{k+1:n}, y_{k+1:n}, z_{k+1:n}) = 1$$

and such that for all $k \in \{1, \dots, n-1\}$ it holds

$$\gamma_t^k(\,\cdot\,|z_{k+1:n}) = \sum_{x_{k+1:n}, y_{k+1:n}} \nu_t^{k,x_{k+1:n},y_{k+1:n}}(\,\cdot\,)\alpha_t^k(x_{k+1:n}, y_{k+1:n}, z_{k+1:n}).$$

Indeed, by definition and using the tensorisation property of $\nu_t^{x,y}$ given in Lemma 2.10, it holds

$$\gamma_t(z) = \sum_{x,y \in V} \nu_t^{x,y}(z)\hat{\pi}(x, y).$$

So, using the fact that, according to Lemma 2.10, $\nu_t^{x,y}(z) = \prod_{i=1}^n \nu_t^{x_i,y_i}(z_i)$, we see that

$$\begin{aligned} \sum_{u \in V : u_{k:n} = z_{k:n}} \gamma_t(u) &= \sum_{x,y \in V} \left(\sum_{u \in V : u_{k:n} = z_{k:n}} \nu_t^{x,y}(u) \right) \hat{\pi}(x, y) = \sum_{x,y \in V} \prod_{i=k}^{n} \nu_t^{x_i,y_i}(z_i)\hat{\pi}(x, y) \\ &= \sum_{x_{k:n}, y_{k:n}} \prod_{i=k}^{n} \nu_t^{x_i,y_i}(z_i)\pi^i(x_i, y_i|x_{i+1:n}, y_{i+1:n}) \\ &= \sum_{x_{k+1:n}, y_{k+1:n}} \nu_t^{k,x_{k+1:n},y_{k+1:n}}(z_k) \prod_{i=k+1}^{n} \nu_t^{x_i,y_i}(z_i)\pi^i(x_i, y_i|x_{i+1:n}, y_{i+1:n}). \end{aligned}$$

From this it follows that

$$\begin{aligned}\gamma_t^k(z_k|z_{k+1:n}) &= \frac{\displaystyle\sum_{u\in V:u_{k:n}=z_{k:n}} \gamma_t(u)}{\displaystyle\sum_{u\in V:u_{k+1:n}=z_{k+1:n}} \gamma_t(u)} = \frac{\displaystyle\sum_{x_{k+1:n},y_{k+1:n}} \nu_t^{k,x_{k+1:n},y_{k+1:n}}(z_k)\prod_{i=k+1}^{n} \nu_t^{x_i,y_i}(z_i)\pi^i(x_i,y_i|x_{i+1:n},y_{i+1:n})}{\displaystyle\sum_{x_{k+1:n},y_{k+1:n}}\prod_{i=k+1}^{n} \nu_t^{x_i,y_i}(z_i)\pi^i(x_i,y_i|x_{i+1:n},y_{i+1:n})}\\ &:= \sum_{x_{k+1:n},y_{k+1:n}} \nu_t^{k,x_{k+1:n},y_{k+1:n}}(z_k)\alpha_t^k(x_{k+1:n},y_{k+1:n},z_{k+1:n}),\end{aligned}$$

using obvious notation, from which the claim follows. Similarly, for all $z_n \in V_n$, it holds $\gamma_t^n(z_n) = \nu_t^n(z_n)$.

Now, let us recall the well known disintegration formula for the relative entropy: if $\gamma \in \mathcal{P}(V)$ is absolutely continuous with respect to μ, then it holds

$$H(\gamma|\mu) = H(\gamma^n|\mu^n) + \sum_{k=1}^{n-1}\sum_{z\in V} H(\gamma^k(\,\cdot\,|z_{k+1:n})|\mu^k)\gamma(z). \tag{4.7}$$

Applying (4.7) to γ_t, and the (classical) convexity of the relative entropy, it holds

$$\begin{aligned}H(\gamma_t|\mu) &= H(\gamma_t^n|\mu^n) + \sum_{k=1}^{n-1}\sum_{z\in V} H(\gamma_t^k(\,\cdot\,|z_{k+1:n})|\mu^k)\gamma_t(z)\\ &\leq H(\nu_t^n|\mu^n) + \sum_{k=1}^{n-1}\sum_{z\in V}\sum_{\substack{x_{k+1:n},\\ y_{k+1:n}}} \alpha_t^k(x_{k+1:n},y_{k+1:n},z_{k+1:n})H(\nu_t^{k,x_{k+1:n},y_{k+1:n}}|\mu^k)\gamma_t(z)\end{aligned}$$

Now we deal with each term in the sum separately. Fix $k \in \{1,\ldots,n-1\}$. We have

$$\begin{aligned}&\sum_{z\in V}\sum_{\substack{x_{k+1:n},\\ y_{k+1:n}}} \alpha_t^k(x_{k+1:n},y_{k+1:n},z_{k+1:n})H(\nu_t^{k,x_{k+1:n},y_{k+1:n}}|\mu^k)\gamma_t(z)\\ &= \sum_{z_{k+1:n}}\sum_{\substack{x_{k+1:n},\\ y_{k+1:n}}} \alpha_t^k(x_{k+1:n},y_{k+1:n},z_{k+1:n})H(\nu_t^{k,x_{k+1:n},y_{k+1:n}}|\mu^k) \sum_{\substack{u\in V:\\ u_{k+1:n}=z_{k+1:n}}} \gamma_t(u)\\ &= \sum_{z_{k+1:n}}\sum_{\substack{x_{k+1:n},\\ y_{k+1:n}}} H(\nu_t^{k,x_{k+1:n},y_{k+1:n}}|\mu^k)\prod_{i=k+1}^{n} \nu_t^{x_i,y_i}(z_i)\pi^i(x_i,y_i|x_{i+1:n}y_{i+1:n})\\ &= \sum_{\substack{x_{k+1:n},\\ y_{k+1:n}}} H(\nu_t^{k,x_{k+1:n},y_{k+1:n}}|\mu^k)\prod_{i=k+1}^{n} \pi^i(x_i,y_i|x_{i+1:n}y_{i+1:n})\\ &= \sum_{x,y} H(\nu_t^{k,x_{k+1:n},y_{k+1:n}}|\mu^k)\prod_{i=1}^{n} \pi^i(x_i,y_i|x_{i+1:n}y_{i+1:n})\,.\end{aligned}$$

Therefore,

$$H(\gamma_t|\mu) \leq H(\nu_t^n|\mu^n) + \sum_{k=1}^{n-1}\sum_{x,y} H(\nu_t^{k,x_{k+1:n},y_{k+1:n}}|\mu^k)\hat{\pi}(x,y).$$

Now, applying the assumed displacement convexity inequalities, we get

$$\begin{aligned}
H(\gamma_t|\mu) &\leq (1-t)\left[H(\nu_0^n|\mu^n) + \sum_{k=1}^{n-1}\sum_{x,y} H(\nu_0^k(\,\cdot\,|x_{k+1:n})|\mu^k)\hat{\pi}(x,y)\right] \\
&\quad + t\left[H(\nu_1^n|\mu^n) + \sum_{k=1}^{n-1}\sum_{x,y} H(\nu_1^k(\,\cdot\,|y_{k+1:n})|\mu^k)\hat{\pi}(x,y)\right] \\
&\quad - Ct(1-t)\left[J_2(\pi^n) + \sum_{k=1}^{n-1}\sum_{x,y} R_2(\pi^k(\,\cdot\,|x_{k+1:n}, y_{k+1:n}))\hat{\pi}(x,y)\right] \\
&= (1-t)\left[H(\nu_0^n|\mu^n) + \sum_{k=1}^{n-1}\sum_{x} H(\nu_0^k(\,\cdot\,|x_{k+1:n})|\mu^k)\nu_0(x)\right] \\
&\quad + t\left[H(\nu_1^n|\mu^n) + \sum_{k=1}^{n-1}\sum_{y} H(\nu_1^k(\,\cdot\,|y_{k+1:n})|\mu^k)\nu_1(y)\right] \\
&\quad - Ct(1-t)\left[J_2(\pi^n) + \sum_{k=1}^{n-1}\sum_{x,y} R_2(\pi^k(\,\cdot\,|x_{k+1:n}, y_{k+1:n}))\hat{\pi}(x,y)\right] \\
&\leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - Ct(1-t)(I_2^{(n)}(\hat{\pi}) + \bar{I}_2^{(n)}(\hat{\pi})),
\end{aligned}$$

where the last inequality follows from the disintegration equality (4.7) for the relative entropy and from the disintegration inequality given in Proposition 3.8. □

As an application of Theorem 4.6, we derive the displacement convexity of entropy property on the hypercube.

Corollary 4.8 (Displacement convexity on the hypercube)**.** *Let μ be a probability measure on $\{0,1\}$ and define its n-fold product $\mu^{\otimes n}$ on $\Omega_n = \{0,1\}^n$. For any $\nu_0, \nu_1 \in \mathcal{P}(\Omega_n)$, there exists a $\pi \in \Pi(\nu_0, \nu_1)$ such that for any $t \in [0,1]$,*

$$H(\nu_t^\pi|\mu^{\otimes n}) \leq (1-t)H(\nu_0|\mu^{\otimes n}) + tH(\nu_1|\mu^{\otimes n}) - \frac{t(1-t)}{2}\left(I_2^{(n)}(\pi) + \bar{I}_2^{(n)}(\pi)\right). \tag{4.9}$$

and there exists $\pi \in \Pi(\nu_0, \nu_1)$ such that for any $t \in [0,1]$,

$$H(\nu_t^\pi|\mu^{\otimes n}) \leq (1-t)H(\nu_0|\mu^{\otimes n}) + tH(\nu_1|\mu^{\otimes n}) - 2t(1-t)J_2^{(n)}(\pi). \tag{4.10}$$

Proof. According to Proposition 4.1, for all $\nu_0, \nu_1 \in \mathcal{P}(\{0,1\})$, it holds

$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{t(1-t)}{2}\left(\widetilde{\mathcal{T}}_2(\nu_1|\nu_0) + \widetilde{\mathcal{T}}_2(\nu_0|\nu_1)\right), \qquad \forall t \in [0,1],$$

with $\nu_t = (1-t)\nu_0 + t\nu_1$. It is not difficult to check that the coupling π defined by (3.5) is optimal for both $\widetilde{\mathcal{T}}_2(\nu_1|\nu_0)$ and $\widetilde{\mathcal{T}}_2(\nu_0|\nu_1)$. Since on the two-point space $\nu_t = \nu_t^\pi$ is independent of π, the preceding inequality can be rewritten as follows:

$$H(\nu_t^\pi|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - \frac{t(1-t)}{2}\left(I_2(\pi) + \bar{I}_2(\pi)\right), \qquad \forall t \in [0,1].$$

Therefore, we are in a position to apply Theorem 4.6, and to conclude that $\mu^{\otimes n}$ verifies the announced displacement convexity property (4.9).

Similarly, by Lemma 3.3, the displacement convexity property (4.3) ensures that

$$H(\nu_t^\pi|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - 2t(1-t)J_2(\pi), \qquad \forall t \in [0,1].$$

The result then follows from Theorem 4.6. □

Let π be a coupling of $\nu_0, \nu_1 \in \mathcal{P}(\Omega_n)$. By the Cauchy-Schwarz inequality, we have

$$\begin{aligned} J_2^{(n)}(\pi) &= \sum_{i=1}^n \left(\sum_{x,y\in\Omega_n} \mathbb{1}_{x_i \neq y_i} \pi(x,y) \right)^2 \geq \frac{1}{n} \left(\sum_{x,y\in\Omega_n} \sum_{i=1}^n \mathbb{1}_{x_i\neq y_i} \pi(x,y) \right)^2 = \frac{1}{n} \left(\sum_{x,y\in\Omega_n} d(x,y)\pi(x,y) \right)^2 \\ &\geq \frac{1}{n} W_1(\nu_1,\nu_0)^2. \end{aligned}$$

We immediately deduce from Corollary 4.8 the following weaker result.

Corollary 4.11. *Let μ be a probability measure on $\{0,1\}$ and define its n-fold product $\mu^{\otimes n}$ on $\Omega_n = \{0,1\}^n$. For any $\nu_0, \nu_1 \in \mathcal{P}(\Omega_n)$, there exists $\pi \in \Pi(\nu_0,\nu_1)$ such that for $t \in [0,1]$,*

$$H(\nu_t^\pi|\mu^{\otimes n}) \leq (1-t)H(\nu_0|\mu^{\otimes n}) + tH(\nu_1|\mu^{\otimes n}) - \frac{2t(1-t)}{n} W_1(\nu_1,\nu_0)^2.$$

The constant $1/n$ encodes, in some sense, the discrete Ricci curvature of the hypercube in accordance with the various definitions of the discrete Ricci curvature (see the introduction).

Remark 4.12. *Since $\widetilde{\mathcal{T}}_2$ is defined as an infimum, one can replace, for free, the term $I_2^{(n)}(\pi)$ by $\widetilde{\mathcal{T}}_2^{(n)}(\nu_1|\nu_0)$ in* (4.9)*. Moreover, if one chooses $\nu_0 = \mu^{\otimes n}$ and uses that $H(\nu_t^\pi|\mu^{\otimes n}) \geq 0$, one easily derives from* (4.9) *the following transport-entropy inequality:*

$$\widetilde{\mathcal{T}}_2^{(n)}(\nu|\mu^{\otimes n}) + \widetilde{\mathcal{T}}_2^{(n)}(\mu^{\otimes n}|\nu) \leq 2H(\nu|\mu^{\otimes n}), \qquad \forall \nu \in \mathcal{P}(\Omega_n).$$

See [17] *for more on such an inequality (on graphs). Note that the above argument is general and that one can always derive from the displacement convexity of the entropy some Talagrand-type transport-entropy inequality.*

5. HWI TYPE INEQUALITIES ON GRAPHS.

As already stated in the introduction, the displacement convexity of entropy property is usually (*i.e.,* in continuous space settings) the strongest property in the following hierarchy:

$$\text{Displacement convexity } \Rightarrow \text{ HWI } \Rightarrow \text{Log Sobolev.}$$

Applying an argument based on the differentiation property of Corollary 2.8, in this section, we derive HWI and log-Sobolev type inequalities from the displacement convexity property.

We shall start with a general statement on product of graphs that allows to obtain symmetric HWI inequality from the displacement convexity property of the entropy. As a consequence, we get a new symmetric HWI inequality on the hypercube that implies a modified log-Sobolev inequality on the hypercube. This modified log-Sobolev inequality also implies, by means of the Central Limit Theorem, the classical log-Sobolev inequality for the standard Gaussian measure, with the optimal constant.

Then we move to another HWI type inequality involving the already mentioned Dirichlet form $\mathcal{E}_\mu(f, \log f)$ based on Equation (2.16) available on complete graph.

5.1. **Symmetric HWI inequality for products of graphs.** The main result of this section is the following abstract symmetric HWI inequality valid on the n-fold product of any graph.

Proposition 5.1 (HWI). *Consider G^n for $G = (V, E)$ any graph and $\mu \in \mathcal{P}(V^n)$. Assume that μ verifies the following displacement convexity inequality: there is some $c > 0$ such that for any $\nu_0, \nu_1 \in \mathcal{P}(V^n)$, there exists a coupling $\pi \in \Pi(\nu_0, \nu_1)$ such that*

$$H(\nu_t^\pi|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - ct(1-t)(I_2^{(n)}(\pi) + \bar{I}_2^{(n)}(\pi)) \qquad \forall t \in [0,1].$$

Then μ verifies

(5.2)

$$H(\nu_0|\mu) \leq H(\nu_1|\mu) + \sqrt{\sum_{x\in V^n}\sum_{i=1}^n \left[\sum_{z\in N_i(x)}\left(\log\frac{\nu_0(x)}{\mu(x)} - \log\frac{\nu_0(z)}{\mu(z)}\right)\right]_+^2 \nu_0(x)}\,\sqrt{I_2^{(n)}(\pi)} - c(I_2^{(n)}(\pi) + \bar{I}_2^{(n)}(\pi)),$$

for the same $\pi \in \Pi(\nu_0, \nu_1)$ as above, where $N_i(x) = \{z \in V^n; d(x,z) = 1 \text{ and } x_i \neq z_i\}$.

The proof of this result is given below. Before proving that, we derive a certain *reinforced* log-Sobolev inequality (see below for a brief justification of the name) in the discrete setting, and as a consequence, the classical Gross' log-Sobolev inequality on the continuous line, with the optimal constant.

Choose $\nu_1 = \mu$ in (5.2) and denote by $f(x) = \nu_0(x)/\mu(x)$. Then, using the elementary inequality $\sqrt{ab} \leq a/(2\varepsilon) + \varepsilon b/2$, $\varepsilon > 0$, we immediately get the following corollary.

Corollary 5.3 (Reinforced log-Sobolev). *Under the same assumptions of Proposition 5.1, for all $f\colon V^n \to (0,\infty)$ with $\mu(f) = 1$, for all $\varepsilon \leq 2c$, it holds that*

$$\mathrm{Ent}_\mu(f) \leq \frac{1}{2\varepsilon}\sum_{x\in V^n}\sum_{i=1}^n\left[\sum_{z\in N_i(x)}(\log f(x) - \log f(z))\right]_+^2 f(x)\mu(x) - (c - \frac{\varepsilon}{2})\widetilde{\mathcal{T}}_2(\mu|f\mu) - c\widetilde{\mathcal{T}}_2(f\mu|\mu). \tag{5.4}$$

Inequality (5.4) can be seen as a reinforcement of a (discrete) modified log-Sobolev inequality. The next corollary deals with the special case of the discrete cube.

Corollary 5.5 (Reinforced log-Sobolev on Ω_n and Gross' Inequality). *Let μ be a Bernoulli measure on $\{0,1\}$. Then, for any n and any $f\colon \Omega_n \to (0,\infty)$, it holds*

$$\mathrm{Ent}_{\mu^{\otimes n}}(f) \leq \frac{1}{2}\sum_{x\in\Omega_n}\sum_{i=1}^n [\log f(x) - \log f(\sigma_i(x))]_+^2 f(x)\mu^{\otimes n}(x) - \frac{1}{2}\widetilde{\mathcal{T}}_2(f\mu|\mu), \tag{5.6}$$

where $\sigma_i(x) = (x_1, \ldots, x_{i-1}, 1 - x_i, x_{i+1}, \ldots, x_n)$ is the neighbor of $x = (x_1, \ldots, x_n)$ for which the i-th coordinate differs from that of x.

As a consequence, for any n and any $g\colon \mathbb{R}^n \to \mathbb{R}$ smooth enough, it holds

$$\mathrm{Ent}_{\gamma_n}(e^g) \leq \frac{1}{2}\int |\nabla g|^2 e^g d\gamma_n \tag{5.7}$$

where γ_n is the standard Gaussian measure on $\mathbb{R}^n$, and $|\nabla g|$ is the length of the gradient of g.

Remark 5.8. *Note that the constant $1/2$ in the above log-Sobolev inequality for the standard Gaussian is optimal, see e.g.* [1, Chapter 1].

We proceed with the proofs of Proposition 5.1 and Corollary 5.5.

Proof of Proposition 5.1. The displacement convexity inequality ensures that for all $t \in [0,1]$,

$$H(\nu_0|\mu) \leq H(\nu_1|\mu) - \frac{H(\nu_t|\mu) - H(\nu_0|\mu)}{t} - c(I_2^{(n)}(\pi) + \bar{I}_2^{(n)}(\pi)).$$

As t goes to 0, this yields

$$H(\nu_0|\mu) \leq H(\nu_1|\mu) - \frac{\partial}{\partial t} H(\nu_t^\pi|\mu)_{|t=0} - c(I_2^{(n)}(\pi) + \bar{I}_2^{(n)}(\pi)),$$

where $\pi \in \Pi(\nu_0, \nu_1)$. According to Corollary 2.8, it holds

$$\begin{aligned}
-\frac{\partial}{\partial t} H(\nu_t^\pi|\mu)_{|t=0} &= \sum_{\substack{x,z\in V^n:\\ z\sim x}} \left(\log\frac{\nu_0(x)}{\mu(x)} - \log\frac{\nu_0(z)}{\mu(z)}\right) \sum_{y\in V^n} d(x,y)\frac{|\Gamma(x,z,y)|}{|\Gamma(x,y)|}\pi(x,y)\\
&= \sum_{\substack{x,z\in V^n:\\ z\sim x}} \left(\log\frac{\nu_0(x)}{\mu(x)} - \log\frac{\nu_0(z)}{\mu(z)}\right) \sum_{y\in V^n} d(x,y)\frac{|\Gamma(x,z,y)|}{|\Gamma(x,y)|}\pi(x,y)\\
&\leq \sum_{x\in V^n}\sum_{i=1}^n \left[\sum_{z\in N_i(x)} \left(\log\frac{\nu_0(x)}{\mu(x)} - \log\frac{\nu_0(z)}{\mu(z)}\right)\right]_+ \sum_{y\in V^n} d(x,y)\frac{|\Gamma(x,z,y)|}{|\Gamma(x,y)|}\pi(x,y).
\end{aligned}$$

According to (2.11), by induction on $n \geq 1$, we get that for all $u, y \in V^n$,

$$|\Gamma(u,y)| = \frac{d(u,y)!}{\prod_{j=1}^n d(u_j,y_j)!}\prod_{j=1}^n |\Gamma(u_j,y_j)|.$$

Applying this formula with $u = z \in N_i(x)$ for some $i \in \{1,\dots,n\}$ and $u = x$, we get that for all y such that $z \in [\![x,y]\!]$, it holds

$$\frac{|\Gamma(x,z,y)|}{|\Gamma(x,y)|} = \frac{|\Gamma(z,y)|}{|\Gamma(x,y)|} = \frac{d(z,y)!}{d(x,y)!}\frac{d(x_i,y_i)!}{d(z_i,y_i)!}\frac{|\Gamma(z_i,y_i)|}{|\Gamma(x_i,y_i)|} = \frac{d(x_i,y_i)}{d(x,y)}\frac{|\Gamma(z_i,y_i)|}{|\Gamma(x_i,y_i)|}, \tag{5.9}$$

using that $x_j = z_j$ for all $i \neq j$ and the relations $d(x,y) = 1 + d(z,y)$ and $d(x_i,y_i) = 1 + d(z_i,y_i)$. Therefore, when $z \in N_i(x)$,

$$\sum_{y\in V^n} d(x,y)\frac{|\Gamma(x,z,y)|}{|\Gamma(x,y)|}\pi(x,y) = \sum_{y\in V} d(x_i,y_i)\frac{|\Gamma(x_i,z_i,y_i)|}{|\Gamma(x_i,y_i)|}\pi(x,y) \leq \sum_{y\in V} d(x_i,y_i)\pi(x,y).$$

Plugging this inequality into the expression for $-\frac{\partial}{\partial t}H(\nu_t^\pi|\mu)_{|t=0}$ yields:

$$\begin{aligned}
-\frac{\partial}{\partial t} H(\nu_t^\pi|\mu)_{|t=0} &\leq \sum_{x\in V^n}\sum_{i=1}^n \left[\sum_{z\in N_i(x)} \left(\log\frac{\nu_0(x)}{\mu(x)} - \log\frac{\nu_0(z)}{\mu(z)}\right)\right]_+ \sum_{y\in V^n} d(x_i,y_i)\pi(x,y)\\
&\leq \sum_{x\in V^n}\sum_{i=1}^n \left[\sum_{z\in N_i(x)} \left(\log\frac{\nu_0(x)}{\mu(x)} - \log\frac{\nu_0(z)}{\mu(z)}\right)\right]_+ \sum_{y\in V^n} d(x_i,y_i)\frac{\pi(x,y)}{\nu_0(x)}\,\nu_0(x)\\
&\leq \sqrt{\sum_{x\in V^n}\sum_{i=1}^n \left[\sum_{z\in N_i(x)} \left(\log\frac{\nu_0(x)}{\mu(x)} - \log\frac{\nu_0(z)}{\mu(z)}\right)\right]_+^2 \nu_0(x)}\,\sqrt{I_2^{(n)}(\pi)},
\end{aligned}$$

where the last line follows from the Cauchy-Schwarz inequality. This completes the proof. □

Proof of Corollary 5.5. By Corollary 4.8, Inequality (5.4) holds with $c = 1/2$. Observe that $N_i(x) = \{\sigma_i(x)\}$ where $\sigma_i(x) = (x_1, \dots, x_{i-1}, 1 - x_i, x_{i+1}, \dots, x_n)$ is the neighbor of $x = (x_1, \dots, x_n)$ for which the i-th coordinate differs from that of x. For $\varepsilon = 1$, Corollary 5.3 gives

$$\mathrm{Ent}_{\mu^{\otimes n}}(f) \le \frac{1}{2} \sum_{x \in \Omega_n} \sum_{i=1}^{n} [\log f(x) - \log f(\sigma_i(x))]_+^2 \, f(x)\mu^{\otimes n}(x) - \frac{1}{2}\widetilde{\mathcal{T}}_2(f\mu|\mu),$$

which is the first part of the corollary.

For the second part, we shall apply the Central Limit Theorem. Our starting point is the following modified log-Sobolev inequality on the hypercube:

$$\mathrm{Ent}_{\mu^{\otimes n}}(f) \le \frac{1}{2} \sum_{x \in \Omega_n} \sum_{i=1}^{n} [\log f(x) - \log f(\sigma_i(x))]_+^2 \, f(x)\mu^{\otimes n}(x) \tag{5.10}$$

that holds for all product probability measures on the hypercube $\Omega_n = \{0, 1\}^n$, for all dimensions $n \ge 1$.

First we observe that, by tensorisation of the log-Sobolev inequality (see *e.g.* [1, Chapter 1]), we only need to prove Gross' Inequality (5.7) in dimension one ($n = 1$). Then, thanks to a result by Miclo [37], we know that extremal functions in the log-Sobolev inequality, in dimension one, are monotone. Hence, we can assume that g is monotone and non-decreasing (the case g non-increasing can be treated similarly). Furthermore, for convenience, we first assume that the function $g\colon \mathbb{R} \to \mathbb{R}$ is smooth and compactly supported.

Let μ_p be the Bernoulli probability measure with parameter $p \in [0, 1]$. We apply (5.10) to the function $f = e^{G_n}$, with

$$G_n(x) = g\left(\frac{\sum_{i=1}^n x_i - np}{\sqrt{np(1-p)}}\right), \qquad x \in \Omega_n,$$

so that $\mathrm{Ent}_{\mu_p^{\otimes n}}\left(e^{G_n}\right)$ tends to $\mathrm{Ent}_\gamma(e^g)$ by the Central Limit Theorem. It remains to identify the limit, when n tends to infinity, of the Dirichlet form (the first term in the right-hand side of (5.10)). Let $\bar{x}^i y_i$ denote the vector $(x_1, \dots, x_{i-1}, y_i, x_{i+1}, \dots, x_n)$. Then,

$$\begin{aligned}\sum_{x_i \in \{0,1\}} [G_n(x) - G_n(\sigma_i(x))]_+^2 e^{G_n(x)} \mu_p(x_i) &= p[G_n(\bar{x}^i 1) - G_n(\bar{x}^i 0)]_+^2 e^{G_n(\bar{x}^i 1)} \\ &\quad + (1-p)[G_n(\bar{x}^i 0) - G_n(\bar{x}^i 1)]_+^2 e^{G_n(\bar{x}^i 0)}.\end{aligned}$$

Now, since

$$\begin{aligned}\frac{\sum_{i=1}^n x_i - np}{\sqrt{np(1-p)}} - \frac{\sum_{j \ne i} x_j - (n-1)p}{\sqrt{(n-1)p(1-p)}} &= \frac{x_i}{\sqrt{np(1-p)}} + \frac{1}{\sqrt{p(1-p)}} \sum_{j \ne i} x_j \left(\frac{1}{\sqrt{n}} - \frac{1}{\sqrt{n-1}}\right) \\ &\quad + \frac{p}{\sqrt{p(1-p)}} \left(\sqrt{n} - \sqrt{n-1}\right) \\ &= \frac{x_i}{\sqrt{np(1-p)}} - \frac{\sum_{j \ne i} x_j}{\sqrt{p(1-p)}\left(\sqrt{n} + \sqrt{n-1}\right)\sqrt{n}\sqrt{n-1}} \\ &\quad + \frac{p}{\sqrt{p(1-p)}\left(\sqrt{n} + \sqrt{n-1}\right)} = O\left(\frac{1}{\sqrt{n}}\right),\end{aligned}$$

by a Taylor Expansion, we have

$$G_n(\bar{x}^i 1) - G_n(\bar{x}^i 0) = \frac{1}{\sqrt{np(1-p)}}\, g'\left(\frac{\sum_{j\neq i} x_j - p(n-1)}{\sqrt{(n-1)p(1-p)}}\right) + O\left(\frac{1}{n}\right).$$

Setting $y_i(x) = \dfrac{\sum_{j\neq i} x_j - p(n-1)}{\sqrt{(n-1)p(1-p)}}$, it follows that

$$\sum_{x_i\in\{0,1\}} [G_n(x) - G_n(\sigma_i(x))]_+^2 e^{G_n(x)}\, \mu_p(x_i) = \frac{g'\,(y_i(x))^2\, e^{g(y_i(x))}}{n(1-p)} + O\left(\frac{1}{n^{3/2}}\right).$$

Now, since all $y_i(x)$'s have the same law under $\mu_p^{\otimes n}$, it follows that

$$\sum_{x\in\Omega_n}\sum_{i=1}^n [G_n(x) - G_n(\sigma_i(x))]_+^2 e^{G_n(x)}\, \mu_p^{\otimes n}(x) = \sum_{x\in\Omega_n} \frac{g'\,(y_1(x))^2\, e^{g(y_1(x))}}{1-p} \mu_p^{\otimes n}(x) + O\left(\frac{1}{\sqrt{n}}\right).$$

The desired result follows by the Central Limit Theorem, then optimizing over all $p \in [0,1]$, and finally by a standard density argument. This ends the proof. □

5.2. **Complete graph.** Combining the differentiation property (2.16) together with the displacement convexity on the complete graph of Proposition 4.1, we shall prove the following result.

Proposition 5.11 (HWI type inequality on the complete graph)**.** *Let $\mu \equiv 1/n$ be the uniform measure on the complete graph K_n. Then, for any $f\colon V(K_n) \to (0,\infty)$ with $\int f d\mu = 1$, it holds*

$$\mathrm{Ent}_\mu(f) \leq \mathcal{E}_\mu(f, \log f) - \frac{1}{2}\left(\widetilde{\mathcal{T}}_2(\mu|f\mu) + \widetilde{\mathcal{T}}_2(f\mu|\mu)\right),$$

where

$$\mathcal{E}_\mu(f, \log f) := \frac{1}{2}\sum_{x,y\in K_n} (f(y) - f(x))(\log f(y) - \log f(x))\mu(x)\mu(y)$$

corresponds to the Dirichlet form associated to the Markov chain on K_n that jumps uniformly at random from any vertex to any vertex (i.e. with transition probabilities $K(x,y) = \mu(y) = 1/n$, for any x, $y \in V(K_n)$).

Proof. We follow the same line of proof as in Proposition 5.1. Fix $f\colon V(K_n) \to (0,\infty)$ with $\int f d\mu = 1$. By Proposition 4.1, applied to $\nu_1 = \mu$ (which implies that $H(\nu_1|\mu) = 0$) and $\nu_0 = f\mu$, we have

$$H(\nu_t|\mu) \leq (1-t)H(\nu_0|\mu) - \frac{t(1-t)}{2}\left(\widetilde{\mathcal{T}}_2(\nu_1|\nu_0) + \widetilde{\mathcal{T}}_2(\nu_0|\nu_1)\right)$$

where $\nu_t = (1-t)\nu_0 + t\nu_1$. Hence, as t goes to 0, we get

$$\int f \log f d\mu = H(\nu_0|\mu) \leq -\frac{\partial}{\partial t} H(\nu_t|\mu)_{|_{t=0}} - \frac{1}{2}\left(\widetilde{\mathcal{T}}_2(\nu_1|\nu_0) + \widetilde{\mathcal{T}}_2(\nu_0|\nu_1)\right).$$

The expected result follows from (2.16). □

In the case of the two-point space, one can deal with any Bernoulli measure (not only the uniform one as in the case of the complete graph).

Proposition 5.12 (HWI for the two-point space)**.** *Let μ be a Bernoulli-p, $p \in (0,1)$ measure on the two-point space $\Omega_1 = \{0,1\}$. Then, for any $f\colon \Omega_1 \to (0,\infty)$ with $\mu(f) = 1$, it holds*

$$\mathrm{Ent}_\mu(f) \leq \mathcal{E}_\mu(f, \log f) - \frac{1}{2}\left(\widetilde{\mathcal{T}}_2(\mu|f\mu) + \widetilde{\mathcal{T}}_2(f\mu|\mu)\right)$$

where,

$$\mathcal{E}_\mu(f, \log f) = p(1-p)(f(1) - f(0))(\log f(1) - \log f(0)).$$

Proof. Reasoning as above, Proposition 4.1, applied to $\nu_1 = \mu$ and $\nu_0 = f\mu$, implies

$$\mathrm{Ent}_\mu(f) \leq -\frac{\partial}{\partial t} H(\nu_t|\mu)_{|_{t=0}} - \frac{1}{2}\left(\widetilde{\mathcal{T}}_2(\mu|f\mu) + \widetilde{\mathcal{T}}_2(f\mu|\mu)\right),$$

where $\nu_t = (1-t)f\mu + t\mu$. Set $q = 1-p$. Since $H(\nu_t|\mu) = [(1-t)f(0)q + tq]\log[(1-t)f(0) + t] + [(1-t)f(1)p + tp]\log[(1-t)f(1) + t]$, it immediately follows that

$$\begin{aligned}\frac{\partial}{\partial t} H(\nu_t|\mu)_{|_{t=0}} &= q(1-f(0))\log f(0) + q(1-f(0)) + p(1-f(1))\log f(1) + p(1-f(1)) \\ &= q(1-f(0))\log f(0) + p(1-f(1))\log f(1)\end{aligned}$$

where the second equality follows from the fact that $p + q = 1 = \mu(f) = qf(0) + pf(1)$. Using again that $1 = qf(0) + pf(1)$, we observe that

$$q(1-f(0))\log f(0) = pq(f(1) - f(0))\log f(0)$$

and

$$p(1-f(1))\log f(1) = -pq(f(1) - f(0)\log f(1),$$

from which the expected result follows. □

6. Prekopa-Leindler type inequality

In this section we show by a duality argument that the displacement convexity property implies a discrete version of the Prekopa-Leindler inequality. (This argument was originally done by J. Lehec [27] in the context of Brascamp-Lieb inequalities.) Then we show that this Prekopa-Leindler inequality allows to recover the discrete modified log-Sobolev inequality (5.10) and a weak version of the transport entropy inequality of Remark 4.12.

Let us first recall the statement of the usual Prekopa-Leindler inequality.

Theorem 6.1 (Prekopa-Leindler [43, 44, 28])**.** *Let $n \in \mathbb{N}^*$ and $t \in [0,1]$. For all triples (f,g,h) of measurable functions on $\mathbb{R}^n$ such that*

$$h((1-t)x + ty) \geq (1-t)f(x) + tg(y), \qquad \forall x,y \in \mathbb{R}^n,$$

it holds

$$\int e^{h(z)}\,dz \geq \left(\int e^{f(x)}\,dx\right)^{1-t}\left(\int e^{g(y)}\,dy\right)^{t}.$$

Using the identity (with $\|\cdot\|$ denoting the Euclidean norm),

$$\frac{1}{2}\|(1-t)x + ty\|_2^2 = (1-t)\frac{\|x\|_2^2}{2} + t\frac{\|y\|_2^2}{2} - t(1-t)\frac{\|x-y\|_2^2}{2}, \qquad x,y \in \mathbb{R}^n,$$

one can recast, without loss, the preceding result into an inequality for the Gaussian distribution.

Theorem 6.2 (Prekopa-Leindler: the Gaussian case)**.** *Let γ_n be the standard normal distribution on $\mathbb{R}^n$ and $t \in [0,1]$. For all triples (f,g,h) of measurable functions on $\mathbb{R}^n$ such that*

$$h((1-t)x+ty) \geq (1-t)f(x) + tg(y) - \frac{t(1-t)}{2}\|x-y\|_2^2, \qquad \forall x,y \in \mathbb{R}^n, \tag{6.3}$$

it holds that

$$\int e^{h(z)}\,\gamma_n(dz) \geq \left(\int e^{f(x)}\,\gamma_n(dx)\right)^{1-t}\left(\int e^{g(y)}\,\gamma_n(dy)\right)^{t}.$$

The next result shows that a discrete Prekopa-Leindler inequality can be derived from the displacement convexity property of the relative entropy.

Theorem 6.4 (Prekopa-Leindler (discrete version))**.** *Let $n \in \mathbb{N}^*$, $t \in [0,1]$ and $\mu \in \mathcal{P}(V^n)$. Suppose that μ verifies the following property: for any $\nu_0, \nu_1 \in \mathcal{P}(V^n)$, there exists a coupling $\pi \in \Pi(\nu_0,\nu_1)$ such that*

$$H(\nu_t^\pi|\mu) \leq (1-t)H(\nu_0|\mu) + tH(\nu_1|\mu) - ct(1-t)I_2^{(n)}(\pi). \tag{6.5}$$

If (f,g,h) is a triple of functions on V^n such that: $\forall x \in V^n$, $\forall m \in \mathcal{P}(V^n)$,

$$\iint h(z)\,\nu_t^{x,y}(dz)m(dy) \geq (1-t)f(x) + t\int g(y)\,m(dy) - ct(1-t)\sum_{i=1}^n \left(\int d(x_i,y_i)\,m(dy)\right)^2, \tag{6.6}$$

then it holds

$$\int e^{h(z)}\,\mu(dz) \geq \left(\int e^{f(x)}\,\mu(dx)\right)^{1-t}\left(\int e^{g(y)}\,\mu(dy)\right)^{t}.$$

Proof. Let $n \in \mathbb{N}$, $f,g,h : V^n \mapsto \mathbb{R}$, $\mu \in \mathcal{P}(V^n)$, $t \in [0,1]$ and $c \in (0,\infty)$ satisfying the hypotheses of the theorem. Given $\nu_0,\nu_1 \in \mathcal{P}(V^n)$, let π be such that (6.5) holds and let p be such that $\pi(x,y) = \nu_0(x)p(x,y)$, $x,y \in V^n$.

Then, integrate (6.6) in the variable x with respect to ν_0, with $m(y) = p(x,y)$, so that (recalling (2.4))

$$\int h\,d\nu_t^\pi \geq (1-t)\int f\,d\nu_0 + t\int g\,d\nu_1 - ct(1-t)I_2^{(n)}(\pi).$$

Together with (6.5), we end up with

$$\int h\,d\nu_t^\pi - H(\nu_t^\pi|\mu) \geq (1-t)\left(\int f d\nu_0 - H(\nu_0|\mu)\right) + t\left(\int g\,d\nu_1 - H(\nu_1|\mu)\right).$$

The result follows by optimization, since by duality (for any $\alpha\colon V^n \mapsto \mathbb{R}$),

$$\sup_{m\in\mathcal{P}(V^n)}\left\{\int \alpha\,dm - H(m|\mu)\right\} = \log\int e^\alpha\,d\mu.$$

This ends the proof. □

An immediate corollary is a Prekopa-Leindler inequality on the discrete hypercube.

Corollary 6.7. *Let μ be a probability measure on $\{0,1\}$, $n \in \mathbb{N}^*$ and $t \in [0,1]$. For all triple (f,g,h) verifying (6.6) with $c = 1/2$, it holds*

$$\int e^{h(z)}\,\mu^{\otimes n}(dz) \geq \left(\int e^{f(x)}\,\mu^{\otimes n}(dx)\right)^{1-t}\left(\int e^{g(y)}\,\mu^{\otimes n}(dy)\right)^{t}.$$

It is well known that Talagrand's transport-entropy inequality and the logarithmic Sobolev inequality for the Gaussian measure are both consequences of the Prekopa-Leindler inequality of Theorem 6.2 [4]. Similarly the discrete version of Prekopa Leindler inequality implies the modified logarithmic Sobolev inequality induced by Corollary 5.3 and the transport-entropy inequality associated with the distance $\widetilde{\mathcal{T}}_2$ of Remark 4.12.

Corollary 6.8. *Assume that the following Prekopa-Leindler inequality holds: for all* $t \in (0,1)$, *for all triples of functions* (f, g, h) *on* V^n *such that:* $\forall x \in V^n$, $\forall m \in \mathcal{P}(V^n)$,

$$\iint h(z)\, \nu_t^{x,y}(dz) m(dy) \geq (1-t) f(x) + t \int g(y)\, m(dy) - ct(1-t) \sum_{i=1}^n \left(\int d(x_i, y_i)\, m(dy) \right)^2,$$

it holds that

$$\int e^{h(z)}\, \mu(dz) \geq \left(\int e^{f(x)}\, \mu(dx) \right)^{1-t} \left(\int e^{g(y)}\, \mu(dy) \right)^t.$$

Then one has, for all functions $h \colon V^n \to \mathbb{R}$,

$$\operatorname{Ent}_\mu(e^h) \leq \frac{1}{4c} \sum_{x \in V^n} \sum_{i=1}^n \left[\sum_{z \in N_i(x)} (h(x) - h(z)) \right]_+^2 e^{h(x)} \mu(x).$$

and for all probability measures ν, *absolutly continous with respect to* μ,

$$c\widetilde{\mathcal{T}}_2(\mu|\nu) \leq H(\nu|\mu), \tag{6.9}$$

$$c\widetilde{\mathcal{T}}_2(\nu|\mu) \leq H(\nu|\mu), \tag{6.10}$$

Proof. We first prove the transport-entropy inequalities (6.9) and (6.10). Let k be a function on V^n (necessarily bounded, since V is finite). We apply the discrete Prekopa-Leindler inequality with $h = 0$, $g = -(1-t)k$ and $f = tQk$, with Qk defined so that the condition (6.6) holds: for all $x \in V^n$,

$$Qk(x) = \inf_{m \in \mathcal{P}(V^n)} \left\{ \int k(y)\, m(dy) + c \sum_{i=1}^n \left(\int d(x_i, y_i)\, m(dy) \right)^2 \right\}.$$

Therefore, one has for all $t \in (0,1)$,

$$\left(\int e^{tQk} d\mu \right)^{1/t} \left(\int e^{-(1-t)k}\, d\mu \right)^{1/(1-t)} \leq 1.$$

As t goes to 1, we get for all functions k on V^n,

$$\int e^{Qk} d\mu \leq e^{\mu(k)},$$

and this is known to be a dual form of the transport-entropy inequality (6.9) (see [17]). Similarly as t goes to 0, we get for all functions k on V^n,

$$\int e^{-k} d\mu \leq e^{-\mu(Qk)},$$

which is a dual form of the transport-entropy inequality (6.10).

Let us now turn to the proof of the modified discrete logarithmic Sobolev inequality. Fix a bounded function $h : V^n \to \mathbb{R}$ and choose $g = th$ and $f = h + tR_t h$ with $R_t h$ designed so that condition (6.6) holds. Namely, for all $x \in V^n$,

$$R_t h(x) = \inf_m \left\{ \frac{1}{t(1-t)} \left(\iint h(z) \nu_t^{x,y}(dz)\, m(dy) - (1-t)h(x) \right) \right.$$
$$\left. - \frac{t}{1-t} \int h(y)\, m(dy) + c \sum_{i=1}^n \left(\int d(x_i, y_i)\, m(dy) \right)^2 \right\},$$

where the infimum runs over all probability measures $m \in \mathcal{P}(V^n)$. Then the Prekopa-Leindler inequality reads

$$\int e^h d\mu \geq \left(\int e^h e^{tR_t h} d\mu \right)^{1-t} \left(\int e^{th} d\mu \right)^t,$$

which can be rewritten as

$$1 \geq \left(\int e^{tR_t h} d\mu_h \right)^{1/t} \left(\int e^{(t-1)h} d\mu_h \right)^{1/(1-t)},$$

with $d\mu_h = \frac{e^h}{\int e^h\, d\mu}\, d\mu$. Letting t go to 0, we easily deduce (leaving some details to the reader) that,

$$\int (\liminf_{t\to 0} R_t h) e^h d\mu \leq \int e^h d\mu \log \int e^h d\mu\,.$$

This can equivalently be written as

$$\mathrm{Ent}_\mu(e^h) \leq \int (h - \liminf_{t\to 0} R_t h) e^h d\mu.$$

We conclude using the following claim.

Claim 6.11. *For all $x \in \mathbb{R}$, we have*

$$h(x) - \liminf_{t\to 0} R_t h(x) \leq \frac{1}{4c} \sum_{i=1}^n \left[\sum_{z \in N_i(x)} (h(x) - h(z)) \right]_+^2 .$$

□

Proof of Claim 6.11. By a Taylor expansion and by Proposition 2.7, for all $x, y \in V^n$,

$$\int h(z) \nu_t^{x,y}(dz) = \nu_t^{x,y}(h) = \nu_0^{x,y}(h) + t d(x,y) \nu_0^{x,y}(\nabla^{x,y} h) + o(t) = h(x) + t d(x,y) \nabla^{x,y} h(x) + o(t),$$

with the quantity $o(t)$ independent of y since h is bounded. Now, from the definition of the sets $N_i(x)$, $i \in \{1, \ldots, n\}$ and using the identity (5.9), one has

$$\nabla^{x,y} h(x) = \frac{1}{|\Gamma(x,y)|} \sum_{\gamma \in \Gamma(x,y)} (h(\gamma_+(x)) - h(x)) = \sum_{z \in V_n, z \sim x} (h(z) - h(x)) \frac{|\Gamma(x,z,y)|}{|\Gamma(x,y)|}$$
$$= \sum_{i=1}^n \sum_{z \in N_i(x)} (h(z) - h(x)) \frac{d(x_i, y_i) |\Gamma(x_i, z_i, y_i)|}{d(x,y) |\Gamma(x_i, y_i)|}.$$

Therefore

$$\begin{aligned}
h(x) - R_t h(x) &= \sup_m \Bigg\{ \int \sum_{i=1}^n \sum_{z \in N_i(x)} (h(x) - h(z))\, d(x_i, y_i) \frac{|\Gamma(x_i, z_i, y_i)|}{|\Gamma(x_i, y_i)|} m(dy) \\
&\qquad - c \sum_{i=1}^n \left(\int d(x_i, y_i)\, m(dy) \right)^2 \Bigg\} + o(1) \\
&\le \sum_{i=1}^n \sup_m \left\{ \left[\sum_{z \in N_i(x)} (h(x) - h(z)) \right]_+ \int d(x_i, y_i) m(dy) - c \left(\int d(x_i, y_i)\, m(dy) \right)^2 \right\} + o(1) \\
&\le \sum_{i=1}^n \sup_{v \ge 0} \left\{ v \left[\sum_{z \in N_i(x)} (h(x) - h(z)) \right]_+ - cv^2 \right\} + o(1) = \frac{1}{4c} \sum_{i=1}^n \left[\sum_{z \in N_i(x)} (h(x) - h(z)) \right]_+^2 + o(1).
\end{aligned}$$

The claim follows by letting t go to 0. □

Acknowledgements. The authors thank Erwan Hillion for informing them of his independent work [20] and his thesis [19]. Thanks also to Nicola Gigli for informing the authors of recent preprints [12, 15]. The French authors thank the hospitality of Georgia Institute of Technology, Atlanta, Georgia. The last author is grateful to his French collaborators for bringing him back to this topic, and to the Université Paris Est Marne la Vallée - Laboratoire d'Analyse et de Mathématiques Appliquées – for their generosity and hospitality in hosting him.

References

[1] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2000. With a preface by Dominique Bakry and Michel Ledoux.

[2] D. Bakry. L'hypercontractivité et son utilisation en théorie des semigroupes. In *Lectures on probability theory (Saint-Flour, 1992)*, volume 1581 of *Lecture Notes in Math.*, pages 1–114. Springer, Berlin, 1994.

[3] S. G. Bobkov and M. Ledoux. On modified logarithmic Sobolev inequalities for Bernoulli and Poisson measures. *J. Funct. Anal.*, 156(2):347–365, 1998.

[4] S. G. Bobkov and M. Ledoux. From Brunn-Minkowski to Brascamp-Lieb and to logarithmic Sobolev inequalities. *Geom. Funct. Anal.*, 10(5):1028–1052, 2000.

[5] S.G. Bobkov and P. Tetali. Modified logarithmic Sobolev inequalities in discrete settings. *J. Theoret. Probab.*, 19(2):289–336, 2006.

[6] A.I. Bonciocat and K.T. Sturm. Mass transportation and rough curvature bounds for discrete spaces. *J. Funct. Anal.*, 256(9):2944–2966, 2009.

[7] P. Caputo, P. Dai Pra, and G. Posta. Convex entropy decay via the Bochner-Bakry-Emery approach. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(3):734–753, 2009.

[8] D. Cordero-Erausquin, R. J. McCann, and M. Schmuckenschläger. A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.*, 146(2):219–257, 2001.

[9] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318, 1967.

[10] P. Dai Pra, A. M. Paganoni, and G. Posta. Entropy inequalities for unbounded spin systems. *Ann. Probab.*, 30(4):1959–1976, 2002.

[11] A. Dembo. Information inequalities and concentration of measure. *Ann. Probab.*, 25(2):927–939, 1997.

[12] M. Erbar. Gradient flows of the entropy for jump processes. Preprint Available at http://www-wt.iam.uni-bonn.de/ erbar/download/Paper4.pdf, 2012.

[13] M. Erbar and J. Maas. Ricci curvature of finite Markov chains via convexity of the entropy. Preprint Available at arXiv:1111.2687, 2012.

[14] F. Gao and J. Quastel. Exponential decay of entropy in the random transposition and Bernoulli-Laplace models. *Ann. Appl. Probab.*, 13(4):1591–1600, 2003.

[15] N. Gigli and J. Maas. Gromov-hausdorff convergence of discrete transportation metrics. Preprint Available at http://arxiv.org/abs/1207.6501, 2012.

[16] S. Goel. Modified logarithmic Sobolev inequalities for some models of random walk. *Stochastic Process. Appl.*, 114(1):51–79, 2004.

[17] N. Gozlan, C. Roberto, P.M. Samson, and P. Tetali. Transport-entropy inequalities in discrete settings. In preparation, 2012.

[18] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. Math.*, 97(4):1061–1083, 1975.

[19] E. Hillion. Analyse et géométrie dans les espaces métriques mesurés : inégalités de Borell-Brascamp-Lieb et conjecture de Olkin-Shepp. PhD thesis, 2010.

[20] E. Hillion. Contraction of measures on graphs. Preprint, 2012.

[21] E. Hillion, O. Johnson, and Y. Yu. Translation of probability measures on Z. Preprint, 2010.

[22] O. Johnson. Log-concavity and the maximum entropy property of the Poisson distribution. *Stochastic Process. Appl.*, 117(6):791–802, 2007.

[23] A. Joulin. A new Poisson-type deviation inequality for Markov jump processes with positive Wasserstein curvature. *Bernoulli*, 15(2):532–549, 2009.

[24] H. Knothe. Contributions to the theory of convex bodies. *Michigan Math. J.*, 4:39–52, 1957.

[25] S. Kullback. Lower bound for discrimination information in terms of variation. *IEEE Trans. Information Theory*, 4:126127, 1967.

[26] J. Lehec. Private communication. 2012.

[27] J. Lehec. Representation formula for the entropy and functional inequalities. To appear in Annales de l'IHP; available at http://arxiv.org/abs/1006.3028., 2012.

[28] L. Leindler. On a certain converse of Hölder's inequality. In *Linear operators and approximation (Proc. Conf., Oberwolfach, 1971)*, pages 182–184. Internat. Ser. Numer. Math., Vol. 20. Birkhäuser, Basel, 1972.

[29] C. Léonard. Private communication. 2012.

[30] Y. Lin and S.-T. Yau. Ricci curvature and eigenvalue estimate on locally finite graphs. *Math. Res. Lett.*, 17(2):343–356, 2010.

[31] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. of Math. (2)*, 169(3):903–991, 2009.

[32] J. Maas. Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.*, 261(8):2250–2292, 2011.

[33] K. Marton. A measure concentration inequality for contracting Markov chains. *Geom. Funct. Anal.*, 6(3):556–571, 1996.

[34] K. Marton. Erratum to: "A measure concentration inequality for contracting Markov chains" [Geom. Funct. Anal. **6** (1996), no. 3, 556–571; MR1392329 (97g:60082)]. *Geom. Funct. Anal.*, 7(3):609–613, 1997.

[35] K. Marton. On a measure concentration inequality of Talagrand for dependent random variables. *Preprint*, 1999.

[36] R. J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128(1):153–179, 1997.

[37] L. Miclo. Monotonicity of the extremal functions for one-dimensional inequalities of logarithmic Sobolev type. In *Séminaire de Probabilités XLII*, volume 1979 of *Lecture Notes in Math.*, pages 103–130. Springer, Berlin, 2009.

[38] A. Mielke. Geodesic convexity of the relative entropy in reversible Markov chains. To appear in Calc. Var. Part. Diff. Equ., 2012.

[39] Y. Ollivier. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.*, 256(3):810–864, 2009.

[40] Y. Ollivier and C. Villani. A curved Brunn–Minkowski inequality on the discrete hypercube; Or: What is the Ricci curvature of the discrete hypercube? To appear in SIAM J. Discrete Math., 2012.

[41] F. Otto and C. Villani. Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.*, 173(2):361–400, 2000.

[42] M. S. Pinsker. *Information and information stability of random variables and processes*. Translated and edited by Amiel Feinstein. Holden-Day Inc., San Francisco, Calif., 1964.

[43] A. Prékopa. Logarithmic concave measures with application to stochastic programming. *Acta Sci. Math. (Szeged)*, 32:301–316, 1971.

[44] A. Prékopa. On logarithmic concave measures and functions. *Acta Sci. Math. (Szeged)*, 34:335–343, 1973.

[45] M-K. von Renesse and K.T. Sturm. Transport inequalities, gradient estimates, entropy, and Ricci curvature. *Comm. Pure Appl. Math.*, 58(7):923–940, 2005.

[46] C. Roberto. Inégalités de Hardy et de Sobolev logarithmiques. PhD thesis, 2001.

[47] M. Rosenblatt. Remarks on a multivariate transformation. *Ann. Math. Statistics*, 23:470–472, 1952.
[48] L. Saloff-Coste. Lectures on finite Markov chains. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, pages 301–413. Springer, Berlin, 1997.
[49] M. Sammer and P. Tetali. Concentration on the discrete torus using transportation. *Combin. Probab. Comput.*, 18(5):835–860, 2009.
[50] M. D. Sammer. *Aspects of mass transportation in discrete concentration inequalities*. ProQuest LLC, Ann Arbor, MI, 2005. Thesis (Ph.D.)–Georgia Institute of Technology.
[51] P.M. Samson. Concentration of measure inequalities for Markov chains and Φ-mixing processes. *Ann. Probab.*, 28(1):416–461, 2000.
[52] P.M. Samson. Infimum-convolution description of concentration properties of product probability measures, with applications. *Ann. Inst. H. Poincaré Probab. Statist.*, 43(3):321–338, 2007.
[53] K.T. Sturm. Convex functionals of probability measures and nonlinear diffusions on manifolds. *J. Math. Pures Appl. (9)*, 84(2):149–168, 2005.
[54] K.T. Sturm. On the geometry of metric measure spaces. I. *Acta Math.*, 196(1):65–131, 2006.
[55] K.T. Sturm. On the geometry of metric measure spaces. II. *Acta Math.*, 196(1):133–177, 2006.
[56] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.
[57] C. Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.

Université Paris Est Marne la Vallée - Laboratoire d'Analyse et de Mathématiques Appliquées (UMR CNRS 8050), 5 bd Descartes, 77454 Marne la Vallée Cedex 2, France

Université Paris Ouest Nanterre La Défense - Modal'X, 200 avenue de la République 92000 Nanterre, France

School of Mathematics & School of Computer Science, Georgia Institute of Technology, Atlanta, GA 30332
E-mail address: `nathael.gozlan@univ-mlv.fr, cyril.roberto@math.cnrs.fr, paul-marie.samson@univ-mlv.fr, tetali@math.gatech.edu`