

Contraction of Measures on Graphs

Erwan Hillion

Received: 18 September 2012 / Accepted: 14 January 2014 / Published online: 1 February 2014
© Springer Science+Business Media Dordrecht 2014

Abstract Given a finitely supported probability measure μ on a connected graph G , we construct a family of probability measures interpolating the Dirac measure at some given point $o \in G$ and μ . Inspired by Sturm-Lott-Villani theory of Ricci curvature bounds on measured length spaces, we then study the convexity of the entropy functional along such interpolations. Explicit results are given in three canonical cases, when the graph G is either \mathbb{Z}^n , a cube or a tree.

Keywords Ricci curvature · Sturm-Lott-Villani theory · Convexity of entropy · Measure contraction property

Mathematic Subject Classifications (2010) 60B99 · 05C99

1 Introduction

On a metric space (X, d) the W_2 -Wasserstein distance between two probability measures having finite second moments is defined by

$$W_2(\mu, \nu)^2 := \inf_{\pi \in \Pi(\mu, \nu)} \int_{X \times X} d(x, y)^2 d\pi(x, y), \quad (1)$$

where $\Pi(\mu, \nu)$ is the set of couplings of μ and ν , i.e. the set of probability measures on $X \times X$ having μ and ν as first and second marginals. Under weak assumptions, it is possible to prove the existence of a minimizer in Eq. 1, called optimal coupling between μ and ν . The theory of optimal transportation is thoroughly studied in the textbooks [10] and [11] by Villani.

It is possible to go further in the theory of optimal transportation if one makes the assumption that the metric space (X, d) is a length space, i.e. that the distance between two

E. Hillion (✉)

Department of Mathematics, University of Luxembourg, Luxembourg, Grand Duchy of Luxembourg
e-mail: erwan.hillion@uni.lu

points $x, y \in X$ is the infimum of lengths of continuous curves joining x to y , where the length of a curve $\gamma : [0, 1] \rightarrow X$ is defined by

$$L(\gamma) := \sup_{N \geq 1, 0=t_0 < \dots < t_N=1} \sum_{i=0}^{N-1} d(\gamma(t_i), \gamma(t_{i+1})).$$

If this infimum is attained by a certain (possibly non-unique) curve γ , this curve is called a geodesic between x and y . If there exist geodesics between every couple of points $x, y \in X$, the metric space (X, d) is called a geodesic space.

An important result asserts that if (X, d) is a geodesic space, so is the metric space $(\mathcal{P}_2(X), W_2)$ of probability measures on (X, d) with the W_2 distance. It is thus possible to define W_2 -geodesics in this setting. In [8, 9] and [5], Sturm, and independently Lott and Villani, study the behaviour of the entropy functional along W_2 geodesics and use its convexity properties to define a notion of Ricci curvature bounds on the underlying geodesic space (X, d) .

For example, a compact geodesic space (X, d) endowed with a reference Borel positive measure ν is said to satisfy the curvature bound $\text{Ric} \geq K$ if, for every W_2 geodesic $(\mu_t)_{t \in [0,1]}$, we have

$$\text{Ent}_\nu(\mu_t) \leq (1 - t) \text{Ent}_\nu(\mu_0) + t \text{Ent}_\nu(\mu_1) + \frac{Kt(1 - t)}{2} W_2(\mu_0, \mu_1)^2, \tag{2}$$

where the entropy functional Ent_ν is defined by

$$\text{Ent}_\nu(\mu) := \int_X \rho \log(\rho) d\nu$$

when $\mu = \rho\nu$ is absolutely continuous with respect to ν and $\text{Ent}_\nu(\mu) := \infty$ elsewhere, and using the convention $0 \log(0) = 0$. If regularity issues are put aside, Eq. 2 is equivalent to

$$\frac{\partial^2}{\partial t^2} \text{Ent}_\nu(\mu_t) \geq K W_2(\mu_0, \mu_1)^2. \tag{3}$$

On a Riemannian manifold, Sturm-Lott-Villani Ricci curvature bounds are equivalent to the classical definition of Ricci curvature bounds. Furthermore, many interesting geometric and analytic properties, such as Poincaré or log-Sobolev inequalities, hold on a geodesic space satisfying Eq. 2, especially when $K > 0$.

Sturm-Lott-Villani theory does not directly apply when the metric space (X, d) is a graph because, although optimal couplings still exist, the W_2 Wasserstein space associated to a graph is not a geodesic space: in fact, any non-trivial curve in $(\mathcal{P}_2(G), W_2)$ has an infinite length.

It is still an interesting open question to construct, given two probability measures μ_0 and μ_1 on a graph, an interpolating family of measures $(\mu_t)_{t \in [0,1]}$ (seen as a generalization, or a substitute, for a W_2 geodesic) for which the behaviour of the entropy functional reflects geometric properties of the underlying graph.

The purpose of this article is to construct and study such an interpolation in the special case where the initial measure μ_0 is a Dirac measure at a given fixed point $o \in G$. The resulting family $(\mu_t)_{t \in [0,1]}$ thus describes how the final measure $\mu = \mu_1$ is contracted to a Dirac mass. The behaviour of the entropy functional along this contraction can be seen as a discrete version of the measure contraction property studied by Ohta in [6].

The question of using the methods introduced here in order to generalize the results of this article to the case where the initial measure μ_0 is not necessarily a Dirac measure is still open. An answer to this question would provide an interesting generalization of the

Sturm-Lott-Villani theory to the settings of graphs. The main difficulty in this general case is the fact that there could exist more than one coupling between each couple of probability measures. In a work in preparation by the current author, it is shown that, between a couple of finitely supported measures μ_0, μ_1 on \mathbb{Z} , there exists an interpolating family $(\mu_t)_{t \in [0,1]}$ which can be expressed as a mixture of binomial measures along a certain coupling $\pi \in \Pi(\mu_0, \mu_1)$, which in some sense can be seen as discrete version of a Wasserstein geodesic, and along which the entropy functional is convex.

The rest of the article goes as follows. In Section 2 we construct explicitly the contraction $(\mu_t)_{t \in [0,1]}$ given a final measure μ and a base point o on G , which turns out to be expressed as mixture of binomial measures. Section 3 is devoted to the particular case where the graph G is \mathbb{Z} and where μ is supported on \mathbb{Z}_+ . The study of this particular case allows us to introduce some technical tools, in particular a f, g -type decomposition which will be studied in a more general setting in Section 4. Section 5 is about the behaviour of the entropy functional when the graph G is \mathbb{Z}^n (resp. a discrete cube, a tree). It will turn out that the convexity properties of the entropy are similar to those expected in geodesic spaces satisfying a Ricci curvature bound of the type $\text{Ric} \geq 0$ (resp. $\text{Ric} \geq K$ for $K > 0$, $\text{Ric} \geq K$ for $K < 0$).

Remark During the redaction of this article, the author has been made aware of a similar work, see [1]. In this paper, the authors use another type of binomial interpolation of measures on graphs, based on the family of Knothe-Rosenblatt coupling of measures. The study of the entropy along their interpolating families provides interesting non-trivial geometric and analytic results for product spaces. In particular their geometric study of the cube implies Theorem 5.5 of the present article. However, it does not seem possible to use the methods and results of [1] to deduce the Theorems 5.4 and 5.6 of the present article.

2 Construction of the Contraction Family

In this article a graph consists of a collection of points (or vertices) G , and a set of edges E which is a subset of the set of non-ordered couples $x \neq y \in G \times G$. If $(x, y) \in E$, we say that x and y are neighbours and we write $x \sim y$. We assume that each point has a finite number of neighbours.

A curve γ of length l between two points $x, y \in G$ is an application $\gamma : \{0, \dots, l\} \rightarrow G$ such that $\gamma(i) \sim \gamma(i + 1)$ for all $i \in \{0, \dots, l - 1\}$. We assume that every graph considered is connected, i.e. that each couple of points $x, y \in G$ is joined by at least one curve. A geodesic between x and y is a curve of minimal length $l =: d(x, y)$ joining x and y . Geodesics always exist on a connected graph, and the application $(x, y) \mapsto d(x, y)$ defines a distance on G , called the graph distance. We denote by $\Gamma_{x,y}$ the set of geodesics between x and y and by $|\Gamma_{x,y}|$ its cardinality.

Proposition 2.1 *Let a and c be two points of G and $p \in \{0, \dots, d(a, c)\}$. Then*

$$|\Gamma_{a,c}| = \sum_{b \in G: d(a,b)=p} \iota(a, b, c) |\Gamma_{a,b}| |\Gamma_{b,c}|,$$

where $\iota(a, b, c) = 1$ if $d(a, b) + d(b, c) = d(a, c)$ and $\iota(a, b, c) = 0$ elsewhere.

The function ι can be seen as an indicator function of b being on some geodesic joining a to c .

Proof of Proposition 2.1 For every $b \in G$ we define $\mathcal{S}(b) := \Gamma_{a,b} \times \Gamma_{b,c}$. It is clear that these sets are pairwise disjoint. Moreover, we have $|\mathcal{S}(b)| = |\Gamma_{a,b}| |\Gamma_{b,c}|$. To prove the proposition, it thus suffices to show that there exists a bijection ϕ between the set $\Gamma_{a,c}$ and the disjoint union $\sqcup_{b \in \mathcal{G}(a,c,p)} \mathcal{S}(b)$ where $\mathcal{G}(a,c,p) := \{b \in G : \iota(a,b,c) = 1, d(a,b) = p\}$. A natural bijection is $\phi(\gamma) := (\gamma(p), \gamma_1, \gamma_2)$ with $\gamma_1(k) = \gamma(k)$ and $\gamma_2(k) = \gamma(k-p)$. Its inverse is given by the concatenation map $\phi^{-1}(b, \gamma_1, \gamma_2) = \gamma$ with $\gamma(k) = \gamma_1(k)$ for $k \leq p$ and $\gamma(k) = \gamma_2(k-p)$ for $k \geq p$. The fact that $b \in \mathcal{G}(a,c,p)$ guarantees that γ is well-defined (especially at $k = p$) and is a geodesic. \square

We now fix a point $o \in G$, called the base point of the graph, and a finitely supported measure $\mu = \mu_1$. Let μ_0 be the Dirac probability measure at o . There is only one coupling between μ_0 and μ_1 , defined by $\pi(o, x) = \mu(x)$ and by $\pi(z, x) = 0$ if $z \neq o$. Consequently, the W_2 distance is equal to

$$W_2(\mu_0, \mu_1) = \sqrt{\sum_{x \in G} d(o, x)^2 \mu(x)}.$$

Definition 2.2 The contraction of a finitely supported measure μ on G is the family of probability measures $(\mu_t)_{t \in [0,1]}$ defined by

$$\mu_t(x) := \sum_{z \in G} \left(\frac{1}{|\Gamma_{o,z}|} \sum_{\gamma \in \Gamma_{o,z}} \text{bin}_{\gamma,t}(x) \right) \mu(z)$$

where for each geodesic γ of G of length p and each parameter $t \in [0, 1]$, the probability measure $\text{bin}_{\gamma,t}$ on G defined by

$$\text{bin}_{\gamma,t}(x) := \text{bin}_{p,t}(k) = \binom{p}{k} t^k (1-t)^{p-k}$$

if $x = \gamma(k)$ and $\text{bin}_{\gamma,t}(x) := 0$ elsewhere.

Using Proposition 2.1, we can give another formula defining the contraction family:

$$\mu_t(x) = \sum_{z \in G} \iota(o, x, z) \frac{|\Gamma_{o,x}| |\Gamma_{x,z}|}{|\Gamma_{o,z}|} \text{bin}_{d(o,z),t}(d(o, x)) \mu(z). \tag{4}$$

The first steps of the construction of the contraction family are quite natural: we first chose a point $z \in G$ with respect to the measure μ . Then we chose uniformly a geodesic γ between o and z .

The last step of the construction is a bit tricky: in order to interpolate the Dirac measures between o and z along the geodesic γ , we use the binomial family $(\text{bin}_{\gamma,t})_{t \in [0,1]}$, which can thus be seen as the discrete version of the interpolating family $(\delta_{x=\gamma(t)})_{t \in [0,1]}$ which is used in continuous settings.

Several reasons justify the choice of the binomial family:

- (i) We first can see the binomial family as describing the behaviour of a low-temperature random walk on \mathbb{Z} conditioned at $t = 0$ and $t = 1$. In other terms, if $(X_t)_{t \geq 0}$ is the law of the simple random walk on \mathbb{Z} with $X_0 \sim \mu$, then

$$\text{bin}_{n,1-t}(k) = \lim_{\varepsilon \rightarrow 0} \mathbb{P}(X_{\varepsilon t} = k | X_{\varepsilon} = 0).$$

Similarly, if $(X_t)_{t \geq 0}$ is a Markov chain on a finite graph G admitting the normalized counting measure as reversible measure, and such that $X_0 \sim \mu$, we have

$$\mu_{1-t}(k) = \lim_{\varepsilon \rightarrow 0} \mathbb{P}(X_{\varepsilon t} = k | X_{\varepsilon} = 0).$$

This low-temperature behaviour can be linked to recent work by Leonard (see [4]), which constructs optimal couplings and W_2 geodesics from solutions to the so-called Schrödinger problem.

- (ii) Another reason for choosing the binomial family is that it is solution to a discrete version of the transport equation: more precisely, for $n \geq 0$, the family of measures $(\mu_t)_{t \in [0,1]} := (\text{bin}_{n,t})_{t \in [0,1]}$ satisfies:

$$\forall f : \{0, \dots, n\} \rightarrow \mathbb{R}, \quad \frac{\partial}{\partial t} \left(\sum_{k=0}^n f(k) \mu_t(k) \right) = \sum_{k=0}^n (\nabla_{[0,n]} f)(k) \mu_t(k),$$

where the operator $\nabla_{[0,n]}$ is the “spatial derivation on $\{0, \dots, n\}$ ” defined by

$$\nabla_{[0,n]} f(k) = \frac{k}{n} (f(k) - f(k - 1)) + \frac{n - k}{n} (f(k + 1) - f(k)).$$

This can be seen as a generalization of the transport equation $\frac{\partial}{\partial t} \mu(x, t) = -n \frac{\partial}{\partial x} \mu(x, t)$ satisfied in the continuous setting by the family $(\delta(x = nt))_{t \in [0,1]}$ interpolating the Dirac measures at $x = 0$ and $x = n$. A discussion about this discrete transport equation and some properties of the operator $\nabla_{[0,n]}$ can be found in [3].

Although it is not a geodesic for the W_2 distance, the family $(\mu_t)_{t \in [0,1]}$ behaves interestingly for other distances on $\mathcal{P}(G)$. For instance, it is a geodesic for the W_1 distance, as shown by the following:

Proposition 2.3 *The W_1 Wassertein distance $W_1(\mu_0, \mu_t)$ defined by*

$$W_1(\mu_0, \mu_t) := \sum_{x \in G} d(o, x) \mu_t(x)$$

is a linear function of t :

$$W_1(\mu_0, \mu_t) = t W_1(\mu_0, \mu_1).$$

Proof If $\gamma \in \Gamma_{o,z}$ for some $z \in G$, then

$$\sum_{x \in G} d(o, x) \text{bin}_{\gamma,t}(x) = \mathbb{E}_{\text{bin}(d(o,z),t)}[|X|] = td(o, z).$$

We thus have

$$\begin{aligned} \sum_{x \in G} d(o, x) \mu_t(x) &= \sum_{z \in G} \left(\frac{1}{|\Gamma_{o,z}|} \sum_{\gamma \in \Gamma_z} \left(\sum_{x \in G} d(o, x) \text{bin}_{\gamma,t}(x) \right) \right) \mu(z) \\ &= t \sum_{z \in G} d(o, z) \mu(z) = t W_1(\mu_0, \mu_1). \end{aligned}$$

□

Actually, Proposition 2.3 almost holds for the W_2 distance, especially when μ_1 is far from μ_0 . More precisely,

Proposition 2.4

$$W_2(\mu_0, \mu_t)^2 = \sum_{x \in G} d(o, x)^2 \mu_t(x) = t^2 W_2(\mu_0, \mu_1)^2 + t(1 - t)W_1(\mu_0, \mu_1).$$

Proof The proof follows the same lines as the proof of Proposition 2.3. We first compute:

$$\begin{aligned} \sum_{x \in G} \text{bin}_{\gamma,t}(x) d(o, x)^2 &= \mathbb{E}_{\text{bin}(d(o,z),t)} [|X|^2] \\ &= t^2 d(o, z)^2 + t(1 - t)d(o, z), \end{aligned}$$

which implies

$$\sum_{x \in G} d(o, x)^2 \mu_t(x) = t^2 W_2(\mu_0, \mu_1)^2 + t(1 - t)W_1(\mu_0, \mu_1). \quad \square$$

3 The One-Dimensional Case

In this paragraph we focus on the particular case where the graph G is \mathbb{Z} , the base point is $o = 0$ and the final measure $\mu = \mu_1$ is supported on \mathbb{Z}_+ . In this case the contraction family is defined by

$$\forall t \in [0, 1], \forall k \in \mathbb{Z}, \mu_t(k) := \sum_{p \geq 0} \text{bin}_{p,t}(k) \mu(p). \tag{5}$$

The family $(\mu_t)_{t \in [0,1]}$ is called the thinning of μ and has already been widely studied. Interesting references about thinning of measures are Renyi’s article [7] where thinning is defined, [2] where thinning is used to state a “law of small numbers” for measures supported on \mathbb{Z}_+ , and [12] where thinning is used to obtain discrete versions of the entropy power inequality.

We are interested in the behaviour of the function $H(t)$ when the parameter t moves, where

$$H(t) := \text{Ent}_v(\mu_t) := \sum_{x \in G} \mu_t(x) \log(\mu_t(x))$$

is the entropy of μ_t with respect to the counting measure on G .

It is easy to see that $H''(t) = A_t + B_t$, where

$$A_t := \sum_{x \in G} \frac{\partial^2 \mu_t(x)}{\partial t^2} \log(\mu_t(x)), \quad B_t := \sum_{x \in G} \left(\frac{\partial \mu_t(x)}{\partial t} \right)^2 \frac{1}{\mu_t(x)}.$$

We will keep the notations A_t and B_t in the rest of the article.

In this paragraph we give a new proof of the following result due to Johnson and Yu (see [12]):

Theorem 3.1 *Given a contraction family $(\mu_t)_{t \in [0,1]}$ on \mathbb{Z}_+ , the associated entropy function $t \mapsto H(t)$ is convex on $[0, 1]$.*

Proof of Theorem 3.1 We define the families of functions $(f_t)_{t \in [0,1]}$, $(g_t)_{t \in [0,1]}$ by

$$\forall k \geq 0, f_t(k) := \frac{t^k}{k!}, \quad g_t(k) := \sum_{p \geq k} \frac{p!}{(p - k)!} (1 - t)^{p-k} \mu(p)$$

It is clear from Eq. 5 that $\mu_t(k) = f_t(k)g_t(k)$. Moreover (with $f_t(-1) = 0$), we have the following differential equations

$$\frac{\partial f_t(k)}{\partial t} = f_t(k - 1), \quad \frac{\partial g_t(k)}{\partial t} = -g_t(k + 1). \tag{6}$$

From these two equations we deduce

$$\frac{\partial \mu_t(k)}{\partial t} = f_t(k - 1)g_t(k) - f_t(k)g_t(k + 1),$$

$$\frac{\partial^2 \mu_t(k)}{\partial t^2} = f_t(k - 2)g_t(k) - 2f_t(k - 1)g_t(k + 1) + f_t(k)g_t(k + 2),$$

thus:

$$\begin{aligned} A_t &= \sum_{k \geq 0} \frac{\partial^2 \mu_t(k)}{\partial t^2} \log(\mu_t(k)) \\ &= \sum_{k \geq 0} [f_t(k - 2)g_t(k) - 2f_t(k - 1)g_t(k + 1) + f_t(k)g_t(k + 2)] \log(\mu_t(k)) \\ &= \sum_{k \geq 0} f_t(k)g_t(k + 2) \log\left(\frac{\mu_t(k)\mu_t(k + 2)}{\mu_t(k + 1)^2}\right) \\ &= \sum_{k \geq 0} f_t(k)g_t(k + 2) \log\left(\frac{g_t(k)g_t(k + 2)}{g_t(k + 1)^2}\right) \\ &\quad + \sum_{k \geq 0} f_t(k - 2)g_t(k) \log\left(\frac{f_t(k)f_t(k - 2)}{f_t(k - 1)^2}\right). \end{aligned}$$

We now apply the elementary inequality

$$\forall x > 0, \log(x) \geq 1 - \frac{1}{x} \tag{7}$$

to obtain

$$\begin{aligned} A_t &\geq \sum_{k \geq 0} f_t(k)g_t(k + 2) - \frac{g_t(k + 1)^2 f_t(k)}{g_t(k)} \\ &\quad + \sum_{k \geq 0} f_t(k - 2)g_t(k) - \frac{f_t(k - 1)^2 g_t(k)}{f_t(k)} \\ &= 2 \sum_{k \geq 0} f_t(k - 1)g_t(k + 1) - \sum_{k \geq 0} \frac{f_t(k)^2 g_t(k + 1)^2 + f_t(k - 1)^2 g_t(k)^2}{\mu_t(k)} \\ &= - \sum_{k \geq 0} \frac{[f_t(k)g_t(k + 1) - f_t(k - 1)g_t(k)]^2}{f_t(k)g_t(k)} \\ &= -B_t. \end{aligned}$$

We thus have proved that $H''(t) = A_t + B_t \geq 0$. □

4 f, g Decomposition of the Contraction Family

The key to the proof of Theorem 3.1 is the decomposition of μ_t as the product of two functions satisfying simple differential equations. In this section we show that, in the general case, such a decomposition is always possible. More precisely:

Definition 4.1 For every $x \in G$ we define

$$f_t(x) := \frac{t^{d(o,x)}}{d(o,x)!} |\Gamma_{o,x}|, \quad g_t(x) := \frac{\mu_t(x)}{f_t(x)}.$$

It is interesting to notice that $(f_t)_{t \in [0,1]}$ depends only on the graph G and the base point o .

Given some $x \in G$, the function $t \mapsto f_t(x)$ satisfies

$$\frac{\partial f_t(x)}{\partial t} = \frac{d(o,x)}{t} f_t(x). \tag{8}$$

Moreover:

Proposition 4.2 *The family of functions $(f_t)_{t \in [0,1]}$ satisfies*

$$\frac{\partial f_t(x)}{\partial t} = \sum_{y \in \mathcal{E}(x)} f_t(y), \tag{9}$$

where the set $\mathcal{E}(x)$ is defined by

$$\mathcal{E}(x) := \{y \sim x \mid d(o,y) = d(o,x) - 1\}.$$

Proof If $x = o$, Proposition 4.2 is true because $\mathcal{E}(o)$ is empty and $f_t(o)$ is constant. If $x \neq o$, we use Proposition 2.1 with $a = o, c = x$ and $p = d(o,x) - 1$ to find

$$|\Gamma_{o,x}| = \sum_{\substack{y \in G: \\ d(o,y) = d(o,x) - 1}} \iota(o,y,x) |\Gamma_{o,y}| |\Gamma_{y,x}|.$$

But if $\iota(o,y,x) = 1$ and $d(o,y) = d(o,x) - 1$ then $d(y,x) = 1$ and $|\Gamma_{y,x}| = 1$. Furthermore, $\{y \in G : \iota(o,y,x) = 1, d(o,y) = d(o,x) - 1\} = \mathcal{E}(x)$, so we can write

$$|\Gamma_{o,x}| = \sum_{y \in \mathcal{E}(x)} |\Gamma_{o,y}|$$

and Proposition 4.2 follows easily. □

It is a bit more complicated to study the family $(g_t)_{t \in [0,1]}$ because it depends on the measure μ . It is however possible to express it as a mixture of functions similar to f_t :

Proposition 4.3 *The family of functions $(g_t)_{t \in [0,1]}$ can be written*

$$g_t(x) = \sum_w \frac{|\Gamma_{x,w}|}{|\Gamma_{0,w}|} \iota(o,x,w) \frac{d(o,w)!}{d(x,w)!} (1-t)^{d(x,w)} \mu(w).$$

Proof By linearity it suffices to consider the case where μ is a Dirac measure at some point $w \in G$. In this case,

$$\begin{aligned}
 g_t(x) &= \left(\frac{1}{|\Gamma_{0,w}|} \sum_{\substack{\gamma \in \Gamma_w: \\ \gamma(d(o,x))=x}} \text{bin}_{d(o,w),t}(d(o,x)) \right) / \left(\frac{t^{d(o,x)} |\Gamma_{o,x}|}{d(o,x)!} \right) \\
 &= \iota(o,x,w) \left(\frac{|\Gamma_{o,x}| |\Gamma_{x,w}|}{|\Gamma_{o,w}|} \frac{d(o,w)!}{d(o,x)! d(x,w)!} t^{d(o,x)} (1-t)^{d(x,w)} \right) / \left(\frac{t^{d(o,x)} |\Gamma_{o,x}|}{d(o,x)!} \right) \\
 &= \iota(o,x,w) \frac{|\Gamma_{x,w}|}{|\Gamma_{o,w}|} \frac{d(o,w)!}{d(x,w)!} (1-t)^{d(x,w)}. \quad \square
 \end{aligned}$$

It is then possible to state a differential equation satisfied by $(g_t)_{t \in [0,1]}$:

Proposition 4.4 *The family of functions $(g_t)_{t \in [0,1]}$ satisfies*

$$\frac{\partial g_t(x)}{\partial t} = - \sum_{z \in \mathcal{F}(x)} g_t(z), \tag{10}$$

where the set $\mathcal{F}(x)$ is defined by

$$\mathcal{F}(x) := \{z \sim x \mid d(o,z) = d(o,x) + 1\}.$$

Remark There is a duality formula between the collection of sets $(\mathcal{E}(x))_{x \in G}$ and $(\mathcal{F}(x))_{x \in G}$:

$$\mathcal{F}(x) = \{z \sim x \mid x \in \mathcal{E}(z)\}. \tag{11}$$

Proof Proposition 4.4 By linearity again, we can suppose that μ is a Dirac measure at some point $w \in G$, and in this case it is sufficient to show that

$$\forall x \in \text{Supp}(g_t), \quad |\Gamma_{x,w}| = \sum_{z \in \mathcal{F}(x) \cap \text{Supp } g_t} |\Gamma_{z,w}|.$$

By Proposition 2.1 we have

$$|\Gamma_{x,w}| = \sum_{\substack{z \in G: \\ d(x,z)=1}} \iota(x,z,w) |\Gamma_{x,z}| |\Gamma_{z,w}|,$$

and we know that $|\Gamma_{x,z}| = 1$ if $d(x,z) = 1$. Proposition 4.4 will thus be proven if we show that $\forall x \in \text{Supp}(g_t)$,

$$\{z \in G : z \in \mathcal{F}(x) \cap \text{Supp } g_t\} = \{z \in G : d(x,z) = 1, \iota(x,z,w) = 1\}.$$

But μ being a Dirac measure at w implies that $x \in \text{Supp}(g_t) \Leftrightarrow \iota(o,x,w) = 1$. Similarly, we have $z \in \text{Supp}(g_t) \Leftrightarrow \iota(o,z,w) = 1$. Moreover, we have $z \in \mathcal{F}(x) \Leftrightarrow x \in \mathcal{E}(z) \Leftrightarrow (\iota(o,x,z) = 1, d(x,z) = 1)$.

This shows that Proposition 4.4 is true if, for every couple (x,z) such that $d(x,z) = 1$, we have

$$\iota(o,x,z) \iota(o,z,w) = \iota(x,z,w) \iota(o,x,w),$$

and this functional equality is actually true for every couple $(x,z) \in G \times G$: the triangle inequality shows that both sides are equal to 1 if and only if $d(o,w) = d(o,x) + d(x,z) + d(z,w)$. □

We can use the duality in Eq. 11 to state an integration by parts formula:

Proposition 4.5 *Given two finitely supported functions u and v on G ,*

$$\begin{aligned} \sum_{x \in G} \left(\sum_{y \in \mathcal{E}(x)} u(y) \right) v(x) &= \sum_{\substack{s_0, s_1: \\ s_0 \rightarrow s_1}} u(s_0)v(s_1) \\ &= \sum_{x \in G} u(x) \left(\sum_{z \in \mathcal{F}(x)} v(z) \right), \end{aligned}$$

where the notation “ $s_0 \rightarrow s_1$ ” stands for “ $s_0 \in \mathcal{E}(s_1)$ ” (or equivalently “ $s_1 \in \mathcal{F}(s_0)$ ”).

We can similarly state a second-order integration by parts formula:

Proposition 4.6 *Given two finitely supported functions u and v on G ,*

$$\begin{aligned} \sum_{\substack{(s_0, s_1, s_2): \\ s_0 \rightarrow s_1 \rightarrow s_2}} u(s_0)v(s_2) &= \sum_{x \in G} \left(\sum_{\substack{(y', y): \\ y' \rightarrow y \rightarrow x}} u(y') \right) v(x) \\ &= \sum_{x \in G} \left(\sum_{y: y \rightarrow x} u(y) \right) \left(\sum_{z: x \rightarrow z} v(z) \right) \\ &= \sum_{x \in G} u(x) \left(\sum_{\substack{(z, z'): \\ x \rightarrow z \rightarrow z'}} v(z') \right). \end{aligned}$$

We now use the decomposition $\mu_t(x) = f_t(x)g_t(x)$ to study the behaviour of the entropy functional along the contraction of a probability measure μ on G . Let us recall:

$$H''(t) = \sum_{x \in G} \frac{\partial^2 \mu_t(x)}{\partial t^2} \log(\mu_t(x)) + \sum_{x \in G} \left(\frac{\partial \mu_t(x)}{\partial t} \right)^2 \frac{1}{\mu_t(x)} \tag{12}$$

$$=: A_t + B_t. \tag{13}$$

Proposition 4.7 *The first sum in Eq. 12 can be written*

$$\begin{aligned} A_t &= \sum_{\substack{(s_0, s_1, s_2): \\ s_0 \rightarrow s_1 \rightarrow s_2}} f_t(s_0)g_t(s_2) \log \left(\frac{f_t(s_0)f_t(s_2)}{f_t(s_1)^2} \right) \\ &+ \sum_{\substack{(s_0, s_1, s_2): \\ s_0 \rightarrow s_1 \rightarrow s_2}} f_t(s_0)g_t(s_2) \log \left(\frac{g_t(s_0)g_t(s_2)}{g_t(s_1)^2} \right). \end{aligned}$$

Proof Applying twice Proposition 4.2 (resp. Proposition 4.4) yields

$$\frac{\partial^2 f_t(x)}{\partial t^2} = \sum_{y \in \mathcal{E}(x)} \sum_{y' \in \mathcal{E}(y)} f_t(y') = \sum_{\substack{(y', y): \\ y' \rightarrow y \rightarrow x}} f_t(y'), \tag{14}$$

$$\frac{\partial^2 g_t(x)}{\partial t^2} = \sum_{z \in \mathcal{F}(x)} \sum_{z' \in \mathcal{F}(z)} g_t(z') = \sum_{\substack{(z, z'): \\ x \rightarrow z \rightarrow z'}} g_t(z'). \tag{15}$$

Set $h(x) := \log(\mu_t(x))$. Using Eqs. 14, 15 and the first point of Proposition 4.6 we can write:

$$\begin{aligned} \sum_{x \in G} \frac{\partial^2 f_t(x)}{\partial t^2} g_t(x) h(x) &= \sum_{x \in G} \left(\sum_{\substack{(y', y): \\ y' \rightarrow y \rightarrow x}} f_t(y') \right) g_t(x) h(x) \\ &= \sum_{\substack{(s_0, s_1, s_2): \\ s_0 \rightarrow s_1 \rightarrow s_2}} f_t(s_0) g_t(s_2) h(s_2). \end{aligned}$$

Similarly,

$$\sum_{x \in G} \frac{\partial f_t(x)}{\partial t} \frac{\partial g_t(x)}{\partial t} h(x) = - \sum_{\substack{(s_0, s_1, s_2): \\ s_0 \rightarrow s_1 \rightarrow s_2}} f_t(s_0) g_t(s_2) h(s_1)$$

and

$$\sum_{x \in G} \frac{\partial^2 g_t(x)}{\partial t^2} f_t(x) h(x) = \sum_{\substack{(s_0, s_1, s_2): \\ s_0 \rightarrow s_1 \rightarrow s_2}} f_t(s_0) g_t(s_2) h(s_0).$$

We deduce

$$\sum_{x \in G} \frac{\partial^2 \mu_t(x)}{\partial t^2} h(x) = \sum_{\substack{(s_0, s_1, s_2): \\ s_0 \rightarrow s_1 \rightarrow s_2}} f_t(s_0) g_t(s_2) [h(s_0) - 2h(s_1) + h(s_2)]$$

and Proposition 4.7 follows easily. □

It will be convenient to reformulate Proposition 4.7 in the following form:

$$\begin{aligned} A_t &= \sum_{x \in G} \sum_{\substack{(y', y): \\ y' \rightarrow y \rightarrow x}} f_t(y') g_t(x) \log \left(\frac{f_t(x) f_t(y')}{f_t(y)^2} \right) \\ &\quad + \sum_{x \in G} \sum_{\substack{(z, z'): \\ x \rightarrow z \rightarrow z'}} f_t(x) g_t(z') \log \left(\frac{g_t(x) g_t(z')}{g_t(z)^2} \right) \\ &=: \sum_{x \in G} A_{1,t}(x) + \sum_{x \in G} A_{2,t}(x) =: A_{1,t} + A_{2,t}. \end{aligned}$$

We now turn to the second sum in Eq. 12. We first decompose it by writing:

$$\begin{aligned}
 B_t &= \sum_{x \in G} \left(\frac{\partial \mu_t(x)}{\partial t} \right)^2 \frac{1}{\mu_t(x)} \\
 &= \sum_{x \in G} \left(\frac{\partial f_t(x)}{\partial t} \right)^2 \frac{g_t(x)}{f_t(x)} + 2 \sum_{x \in G} \frac{\partial f_t(x)}{\partial t} \frac{\partial g_t(x)}{\partial t} + \sum_{x \in G} \left(\frac{\partial g_t(x)}{\partial t} \right)^2 \frac{f_t(x)}{g_t(x)} \\
 &=: \sum_{x \in G} B_{1,t}(x) + 2 \sum_{x \in G} B_{2,t}(x) + \sum_{x \in G} B_{3,t}(x) \\
 &=: B_{1,t} + 2B_{2,t} + B_{3,t}.
 \end{aligned}$$

The simple form taken by the family $(f_t)_{t \in [0,1]}$ allows us to find simple expressions for $B_{1,t}$ and $B_{2,t}$:

Proposition 4.8 *The first two terms $B_{1,t}$ and $B_{2,t}$ can be expressed in terms of Wasserstein distances:*

$$B_{1,t} = \frac{1-t}{t} W_1(\mu_0, \mu_1) + W_2(\mu_0, \mu_1)^2, \quad B_{2,t} = W_1(\mu_0, \mu_1) - W_2(\mu_0, \mu_1)^2.$$

Proof We use Eq. 8 to write:

$$\begin{aligned}
 \sum_{x \in G} \left(\frac{\partial f_t(x)}{\partial t} \right)^2 \frac{g_t(x)}{f_t(x)} &= \sum_{x \in G} \left(\frac{d(o, x)}{t} f_t(x) \right)^2 \frac{g_t(x)}{f_t(x)} \\
 &= \frac{1}{t^2} \sum_{x \in G} d(o, x)^2 \mu_t(x) \\
 &= \frac{1}{t^2} W_2(\mu_0, \mu_t)^2 \\
 &= W_2(\mu_0, \mu_1)^2 + \frac{1-t}{t} W_1(\mu_0, \mu_1)
 \end{aligned}$$

and

$$\begin{aligned}
 \sum_{x \in G} \frac{\partial f_t(x)}{\partial t} \frac{\partial g_t(x)}{\partial t} &= \frac{1}{t} \sum_{x \in G} d(o, x) f_t(x) \frac{\partial g_t(x)}{\partial t} \\
 &= \frac{1}{t} \sum_{x \in G} d(o, x) \left[\frac{\partial \mu_t(x)}{\partial t} - \frac{\partial f_t(x)}{\partial t} g_t(x) \right] \\
 &= \frac{1}{t} \frac{\partial}{\partial t} \left(\sum_{x \in G} d(o, x) \mu_t(x) \right) - \frac{1}{t^2} \sum_{x \in G} d(o, x)^2 \mu_t(x) \\
 &= \frac{W_1(\mu_0, \mu_1)}{t} - \frac{1}{t^2} \left(t^2 W_2(\mu_0, \mu_1)^2 + t(1-t) W_1(\mu_0, \mu_1) \right) \\
 &= W_1(\mu_0, \mu_1) - W_2(\mu_0, \mu_1)^2. \quad \square
 \end{aligned}$$

Remark It is also possible, using Cauchy-Schwarz inequality, to evaluate the third term $B_{3,t}$ in terms of Wasserstein distances:

$$B_{3,t} \leq \frac{t}{1-t} W_1(\mu_0, \mu_1) + W_2(\mu_0, \mu_1)^2.$$

5 Canonical Examples

In this section we focus on three particular families of graphs: the grid \mathbb{Z}^n , the cube $\{0, 1\}^n$ and trees. For each of these cases, we want to generalize Eq. 3 and we try to find concavity inequalities of the form

$$H''(t) = \frac{\partial^2 \text{Ent}(\mu_t)}{\partial t^2} \geq KW(\mu_0, \mu_1),$$

where W is a distance on the space of probability measures on G , which will be either the W_1 or the W_2 Wasserstein distance, and where the constant K does not depend neither on the final measure μ nor on the parameter t .

In each of these examples we keep the notation introduced hitherto.

5.1 The Grid \mathbb{Z}^n

The first example we study is the graph \mathbb{Z}^n : each point of the graph is a n -uple $(x_1, \dots, x_n) \in \mathbb{Z}^n$ and has got $2n$ neighbours. In order to have simpler notations, it is convenient to chose the origin o as the point $o = (0, \dots, 0)$; it is however easy to convince oneself that, because of the invariance of counting measure by translation, the bound on $H''(t)$ does not depend on the choice of the origin.

The graph distance is the L^1 distance:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

and simple combinatorial arguments give

$$|\Gamma_{x,y}| = \frac{(\sum_{i=1}^n |x_i - y_i|)!}{\prod_{i=1}^n |x_i - y_i|!}.$$

This implies

$$f_t(x) = \frac{t^{d(x,o)}}{d(x,o)!} \frac{(\sum_{i=1}^n |x_i|)!}{\prod_{i=1}^n |x_i|!} = \frac{t^{d(x,o)}}{\prod_{i=1}^n |x_i|!}.$$

We now describe, for a given $x \in G$, the sets $\mathcal{E}(x)$ and $\mathcal{F}(x)$. It turns out that these sets depend on the number of non-zeros coordinates of x . More precisely:

Definition 5.1 For $x = \{x_1, \dots, x_n\} \in \mathbb{Z}^n$ we define the subsets I_x and J_x by

$$I_x := \{i \in \{1, \dots, n\} : x_i \neq 0\},$$

$$J_x := \{i \in \{1, \dots, n\} : x_i \notin \{-1, 0, 1\}\}.$$

For $i \in I_x$ we denote by $u_i(x)$ the vector $(0, \dots, 0, \text{sgn}(x_i), 0, \dots, 0)$.

It follows from the definitions that:

Proposition 5.2 Given a point $x \in G$, the sets $\mathcal{E}(x)$ and $\mathcal{F}(x)$ are described by

$$\mathcal{E}(x) = \{x - u_i(x) : i \in I_x\} \text{ and } \mathcal{F}(x) = \{x + u_i(x) : i \in I_x\}.$$

Furthermore:

Proposition 5.3 *If $z \in \mathcal{F}(x)$ then*

$$\mathcal{F}(z) = \{x + u_i(x) + u_j(x) : i, j \in I_x\}.$$

If $y \in \mathcal{E}(x)$ then

$$\mathcal{E}(y) = \{x - u_i(x) - u_j(x) : i \neq j \in I_x\} \sqcup \{x - 2u_i(x) : i \in J_x\}$$

where \sqcup denotes a disjoint union.

Theorem 5.4 *On \mathbb{Z}^n , the entropy functional is convex along contractions of measures. In other terms, we have*

$$H''(t) \geq 0.$$

Proof From Proposition 5.3 we have

$$\begin{aligned} A_{1,t}(x) &= \sum_{i \neq j \in I_x} f_t(x - u_i(x) - u_j(x))g_t(x) \log \left(\frac{f_t(x - u_i(x) - u_j(x))f_t(x)}{f_t(x - u_i(x))^2} \right) \\ &\quad + \sum_{i \in J_x} f_t(x - 2u_i(x))g_t(x) \log \left(\frac{f_t(x - 2u_i(x))f_t(x)}{f_t(x - u_i(x))^2} \right) \end{aligned}$$

Using the commutativity relation $u_i(x) + u_j(x) = u_j(x) + u_i(x)$ we can transform the first sum:

$$\begin{aligned} A_{1,t}(x) &= 2 \sum_{i < j \in I_x} f_t(x - u_i(x) - u_j(x))g_t(x) \log \left(\frac{f_t(x - u_i(x) - u_j(x))f_t(x)}{f_t(x - u_i(x))f_t(x - u_j(x))} \right) \\ &\quad + \sum_{i \in J_x} f_t(x - 2u_i(x))g_t(x) \log \left(\frac{f_t(x - 2u_i(x))f_t(x)}{f_t(x - u_i(x))^2} \right) \end{aligned}$$

We now remark that if $i \neq j \in I_x$ then

$$f_t(x)f_t(x - u_i(x) - u_j(x)) = f_t(x - u_i(x))f_t(x - u_j(x))$$

and if $i \in J_x$ then

$$\frac{f_t(x - 2u_i(x))f_t(x)}{f_t(x - u_i(x))^2} = \frac{|x_i| - 1}{|x_i|}.$$

Using these formulas and the elementary inequality (7) gives

$$\begin{aligned} A_{1,t}(x) &= \sum_{i \in J_x} f_t(x - 2u_i(x))g_t(x) \log \left(\frac{|x_i| - 1}{|x_i|} \right) \\ &\geq \frac{1}{t^2} \sum_{i \in J_x} \mu_t(x)|x_i|(|x_i| - 1) \left(1 - \frac{|x_i|}{|x_i| - 1} \right) \\ &= -\frac{1}{t^2} \sum_{i \in J_x} \mu_t(x)|x_i|. \end{aligned}$$

Noticing that $J_x \subset I_x$ we have:

$$A_{1,t}(x) \geq -\frac{1}{t^2} \sum_{i \in I_x} \mu_t(x)|x_i| = -\frac{1}{t^2}d(o, x)\mu_t(x).$$

Summing over x and using Proposition 4.8 gives

$$A_{1,t} \geq -\frac{1}{t} W_1(\mu_0, \mu_1) = -(B_{1,t} + B_{2,t}).$$

Using again the fact that $u_i(x) + u_j(x) = u_j(x) + u_i(x)$ and the elementary inequality (7), we can give a lower bound on $A_{2,t}(x)$:

$$\begin{aligned} A_{2,t}(x) &= \sum_{i,j \in I_x} f_t(x) g_t(x + u_i(x) + u_j(x)) \log \left(\frac{g_t(x) g_t(x + u_i(x) + u_j(x))}{g_t(x + u_i(x)) g_t(x + u_j(x))} \right) \\ &= 2 \sum_{i < j \in I_x} f_t(x) g_t(x + u_i(x) + u_j(x)) \log \left(\frac{g_t(x) g_t(x + u_i(x) + u_j(x))}{g_t(x + u_i(x)) g_t(x + u_j(x))} \right) \\ &\quad + \sum_{i \in I_x} f_t(x) g_t(x + 2u_i(x)) \log \left(\frac{g_t(x) g_t(x + 2u_i(x))}{g_t(x + u_i(x))^2} \right) \\ &\geq 2 \sum_{i < j \in I_x} f_t(x) g_t(x + u_i(x) + u_j(x)) - g_t(x + u_i(x)) g_t(x + u_j(x)) \frac{f_t(x)}{g_t(x)} \\ &\quad + \sum_{i \in I_x} f_t(x) g_t(x + 2u_i(x)) - g_t(x + u_i(x))^2 \frac{f_t(x)}{g_t(x)}. \end{aligned}$$

But we have:

$$\begin{aligned} &\sum_{x \in G} \left[2 \sum_{i < j \in I_x} g_t(x + u_i(x)) g_t(x + u_j(x)) \frac{f_t(x)}{g_t(x)} + \sum_{i \in I_x} g_t(x + u_i(x))^2 \frac{f_t(x)}{g_t(x)} \right] \\ &= \sum_{x \in G} \left[\left(\sum_{i \in I_x} g_t(x + u_i(x)) \right)^2 \frac{f_t(x)}{g_t(x)} \right] \\ &= \sum_{x \in G} \left[\left(\frac{\partial g_t(x)}{\partial t} \right)^2 \frac{f_t(x)}{g_t(x)} \right] = B_{3,t}. \end{aligned}$$

On the other hand, using Proposition 4.6, we find

$$\begin{aligned} &\sum_{x \in G} \left[2 \sum_{i < j \in I_x} f_t(x) g_t(x + u_i(x) + u_j(x)) + \sum_{i \in I_x} f_t(x) g_t(x + 2u_i(x)) \right] \\ &= \sum_{x \in G} \left[f_t(x) \sum_{(z,z'): x \rightarrow z \rightarrow z'} g_t(z') \right] \\ &= \sum_{x \in G} \left(\sum_{y: y \rightarrow x} f_t(y) \right) \left(\sum_{z: x \rightarrow z} g_t(z) \right) = -B_{2,t}. \end{aligned}$$

Combining everything finally gives

$$\begin{aligned} H''(t) &= (A_{1,t} + A_{2,t}) + (B_{1,t} + 2B_{2,t} + B_{3,t}) \\ &\geq -(B_{1,t} + B_{2,t}) + (-B_{3,t} - B_{2,t}) + (B_{1,t} + 2B_{2,t} + B_{3,t}) \\ &= 0. \end{aligned}$$

□

5.2 The Cube

The n -dimensional cube can be seen as the vector space $\{0, 1\}^n$ on the field $\mathbb{Z}/2\mathbb{Z}$. We denote by (e_1, \dots, e_n) its canonical basis. The application $\phi : I \mapsto \sum_{i \in I} e_i$ is a bijection between the family of subsets of $\{1, \dots, n\}$ and $\{0, 1\}^n$. We will write $i \in x$ for $i \in \phi^{-1}(x)$.

The set $\{0, 1\}^n$ is turned into a graph G by defining the neighbours of a given $x \in G$ as the n points $x + e_i, i \in \{1 \dots n\}$. It is then easy to compute the distance between two points:

$$d(x, y) = |\{i \in \{1, \dots, n\} : i \in x, i \notin y\}| + |\{i \in \{1, \dots, n\} : i \notin x, i \in y\}|$$

and the number of geodesics between them:

$$|\Gamma_{x,y}| = |d(x, y)|!$$

Consequently the function f_t takes the simple form:

$$f_t(x) = t^{d(x,o)}. \tag{16}$$

Theorem 5.5 *The entropy of a contraction family $(\mu_t)_{t \in [0,1]}$ on a cube satisfies the concavity inequality*

$$H''(t) \leq -\frac{1}{n} W_1(\mu_0, \mu_1)^2$$

Proof Equation 16 implies

$$A_{1,t} = \sum_{y' \rightarrow y \rightarrow x} f_t(y') g_t(x) \log \left(\frac{f_t(y') f_t(x)}{f_t(y)^2} \right) = 0.$$

In order to bound $A_{2,t}$ we use the fact that, for every $x \in G$,

$$\mathcal{E}(x) = \{x + e_i : i \in x\} \text{ and } \mathcal{F}(x) = \{x + e_j : j \notin x\}.$$

We use the description of $\mathcal{F}(x)$ to write

$$A_{2,t}(x) = \sum_{i, j \notin x} f_t(x) g_t(x + e_i + e_j) \log \left(\frac{g_t(x) g_t(x + e_i + e_j)}{g_t(x + e_i)^2} \right).$$

As in \mathbb{Z}^n , we use the property $e_i + e_j = e_j + e_i$ to reorganize the sum and then apply inequality (7) to write

$$\begin{aligned} A_{2,t}(x) &= 2 \sum_{i < j \notin x} f_t(x) g_t(x + e_i + e_j) \log \left(\frac{g_t(x) g_t(x + e_i + e_j)}{g_t(x + e_i) g_t(x + e_j)} \right) \\ &\geq 2 \sum_{i < j \notin x} f_t(x) g_t(x + e_i + e_j) - 2 \sum_{i < j \notin x} g_t(x + e_i) g_t(x + e_j) \frac{f_t(x)}{g_t(x)} \\ &= 2 \sum_{i < j \notin x} f_t(x) g_t(x + e_i + e_j) - \left[\left(\sum_{i \notin x} g_t(z) \right)^2 - \sum_{i \notin x} g_t(z)^2 \right] \frac{f_t(x)}{g_t(x)}. \end{aligned}$$

We now use Proposition 4.6 to write

$$\begin{aligned} \sum_{x \in G} \left[2 \sum_{i < j \notin x} f_t(x) g_t(x + e_i + e_j) \right] &= \sum_{x \in G} f_t(x) \left[\sum_{(z, z') : x \rightarrow z \rightarrow z'} g_t(z') \right] \\ &= \sum_{x \in G} \left(\sum_{y : y \rightarrow x} f_t(y) \right) \left(\sum_{z : x \rightarrow z} g_t(z) \right) \\ &= -B_{2,t}. \end{aligned}$$

Similarly,

$$\sum_{x \in G} \left(\sum_{i \notin x} g_t(z) \right)^2 \frac{f_t(x)}{g_t(x)} = \sum_{x \in G} \left(\frac{\partial g_t(x)}{\partial t} \right)^2 \frac{f_t(x)}{g_t(x)} = B_{3,t}.$$

We bound the remaining term using Cauchy-Schwarz inequality:

$$\sum_{x \in G} \sum_{z \in \mathcal{F}(x)} \mu_t(x) \left(\frac{g_t(z)}{g_t(x)} \right)^2 \geq \frac{1}{\sum_{x \in G} \mu_t(x) |\mathcal{F}(x)|} \left(\sum_{x \in G} \sum_{z \in \mathcal{F}(x)} \frac{g_t(z)}{g_t(x)} \mu_t(x) \right)^2.$$

We have the rough bound

$$\frac{1}{\sum_{x \in G} \mu_t(x) |\mathcal{F}(x)|} \geq \frac{1}{n}, \tag{17}$$

and using Proposition 8 we can calculate

$$\begin{aligned} \sum_{x \in G} \sum_{z \in \mathcal{F}(x)} \frac{g_t(z)}{g_t(x)} \mu_t(x) &= - \sum_{x \in G} f_t(x) \frac{\partial g_t(x)}{\partial t} \\ &= - \frac{\partial}{\partial t} \left(\sum_{x \in G} \mu_t(x) \right) + \sum_{x \in G} g_t(x) \frac{\partial f_t(x)}{\partial t} \\ &= 0 + \frac{1}{t} \sum_{x \in G} d(x, o) \mu_t(x) \\ &= W_1(\mu_0, \mu_1). \end{aligned}$$

We finally have:

$$\begin{aligned} H''(t) &= (A_{1,t} + A_{2,t}) + (B_{1,t} + 2B_{2,t} + B_{3,t}) \\ &\geq 0 + \left(-B_{3,t} - B_{2,t} + \frac{W_1(\mu_0, \mu_1)^2}{n} \right) + (B_{1,t} + 2B_{2,t} + B_{3,t}) \\ &= (B_{1,t} + B_{2,t}) + \frac{W_1(\mu_0, \mu_1)^2}{n} \geq \frac{W_1(\mu_0, \mu_1)^2}{n}. \end{aligned}$$

□

5.3 Trees

In this paragraph we suppose that the graph G is a connected tree, i.e. that every couple of points on G is joined by a unique geodesic. In this case, for any point $x \in G$, we have

$$f_t(x) = \frac{t^{d(x,o)}}{d(x,o)!}$$

and

$$|\mathcal{E}(x)| = 1.$$

Theorem 5.6 *The entropy of a contraction family on a tree G satisfies the concavity inequality*

$$H''(t) \geq \log \left(\sup_{z \in G: z \neq o} |\mathcal{F}(z)| \right) W_2(\mu_0, \mu_1)^2.$$

We can remark that if $z \neq o$, then $|\mathcal{F}(z)| = d_G(z) - 1$, where $d_G(z)$ is the degree of the point z in G , i.e. the number of neighbours of z in G . In particular, if $G = \mathbb{Z}$, we find again that $H''(t) \leq 0$.

Proof of Theorem 5.6 The simple form taken by $f_t(x)$ allows us to bound by below the term $A_{1,t}$:

$$\begin{aligned} A_{1,t} &= \sum_{x \in G} \frac{f_t(x)}{t^2} d(o,x)(d(o,x) - 1) g_t(x) \log \left(\frac{((d(o,x) - 1)!)^2}{d(o,x)!(d(o,x) - 2)!} \right) \\ &= \frac{1}{t^2} \sum_{x \in G} \mu_t(x) d(o,x)(d(o,x) - 1) \log \left(\frac{d(o,x) - 1}{d(o,x)} \right) \\ &\geq \frac{1}{t^2} \sum_{x \in G} \mu_t(x) d(o,x)(d(o,x) - 1) \left(1 - \frac{d(o,x)}{d(o,x) - 1} \right) \\ &= -\frac{W_1(\mu_0, \mu_1)}{t} = -(B_{1,t} + B_{2,t}). \end{aligned}$$

We now want to bound $A_{2,t}$. Given some $x \in G$ and $z \in \mathcal{F}(x)$, we set

$$A_{2,t}(x, z) := \sum_{z' \in \mathcal{F}(z)} g_t(z') \log \left(\frac{g_t(x)}{g_t(z)^2} g_t(z') \right).$$

We apply Jensen's inequality to find

$$A_{2,t}(x, z) \geq \left(\sum_{z' \in \mathcal{F}(z)} g_t(z') \right) \log \left(\frac{g_t(x)}{g_t(z)^2} \frac{\sum_{z' \in \mathcal{F}(z)} g_t(z')}{|\mathcal{F}(z)|} \right)$$

We then separate the term $|\mathcal{F}(z)|$ from the terms in $g_t(x), g_t(z), g_t(z')$ in the logarithm and apply inequality (7):

$$\begin{aligned} A_{2,t}(x, z) &\geq - \left(\sum_{z' \in \mathcal{F}(z)} g_t(z') \right) \log(|\mathcal{F}(z)|) + \left(\sum_{z' \in \mathcal{F}(z)} g_t(z') \right) - \frac{g_t(z)^2}{g_t(x)} \\ &= (1 - \log(|\mathcal{F}(z)|)) \left(\sum_{z' \in \mathcal{F}(z)} g_t(z') \right) - \frac{g_t(z)^2}{g_t(x)}. \end{aligned}$$

Summing over x and z , we thus have

$$\begin{aligned} A_{2,t} &= \sum_{\substack{x, z, z': \\ x \rightarrow z \rightarrow z'}} f_t(x) A_{2,t}(x, z) \\ &\geq \sum_{\substack{x, z, z': \\ x \rightarrow z \rightarrow z'}} (1 - \log(|\mathcal{F}(z)|)) f_t(x) g_t(z') - \sum_{x \in G} \left(\sum_{z: x \rightarrow z} \left(\frac{g_t(z)}{g_t(x)} \right)^2 \right) \mu_t(x) \\ &\geq \left(1 - \log \left(\sup_{z \in G: z \neq \emptyset} |\mathcal{F}(z)| \right) \right) \sum_{\substack{x, z, z': \\ x \rightarrow z \rightarrow z'}} f_t(x) g_t(z') \\ &\quad - \sum_{x \in G} \left(\sum_{z: x \rightarrow z} \left(\frac{g_t(z)}{g_t(x)} \right)^2 \right) \mu_t(x). \end{aligned}$$

Using again Proposition 4.6 yields:

$$\sum_{\substack{x, z, z': \\ x \rightarrow z \rightarrow z'}} f_t(x) g_t(z') = -B_{2,t}.$$

The remaining term can be studied as follows:

$$\begin{aligned} \sum_{x \in G} \mu_t(x) \left(\sum_{z: x \rightarrow z} \left(\frac{g_t(z)}{g_t(x)} \right)^2 \right) &\leq \sum_{(x, z): x \rightarrow z} \left(\frac{g_t(z)}{g_t(x)} \right)^2 \mu_t(x) \\ &\leq \sum_{x \in G} \left(\sum_{z: x \rightarrow z} \frac{g_t(z)}{g_t(x)} \right)^2 \mu_t(x) \\ &= \sum_{x \in G} \left(\frac{\partial g_t(x)}{\partial t} \right)^2 \frac{f_t(x)}{g_t(x)} \\ &= B_{3,t}. \end{aligned}$$

We used the elementary fact that, for $a_1, \dots, a_m \geq 0$, $\sum_{i=1}^m a_i^2 \leq (\sum_{i=1}^m a_i)^2$.
Putting everything together gives:

$$\begin{aligned} -H''(t) &= A_{1,t} + A_{2,t} + B_{1,t} + 2B_{2,t} + B_{3,t} \\ &\geq -(B_{1,t} + B_{2,t}) - (1 - \log \left(\sup_{z \in G: z \neq \emptyset} |\mathcal{F}(z)| \right)) B_{2,t} - B_{3,t} \\ &\quad + B_{1,t} + 2B_{2,t} + B_{3,t} \\ &= \log \left(\sup_{z \in G: z \neq \emptyset} |\mathcal{F}(z)| \right) B_{2,t} \\ &\geq \log \left(\sup_{z \in G: z \neq \emptyset} |\mathcal{F}(z)| \right) W_2(\mu_0, \mu_1)^2. \end{aligned}$$

□

Acknowledgments The author thanks the EPSRC for funding through the project "Information geometry of graphs", EP/I009450/1.

The author also thanks Nathael Gozlan, Cyril Roberto, Paul-Marie Samson and Prasad Tetali for addressing him a preliminary version of their research work (see [1]) and for useful discussions.

References

1. Gozlan, N., Roberto, C., Samson, P.M., Tetali, P.: Displacement convexity of entropy and related inequalities on graphs. *Probab. Theory Relat. Fields* (2014)
2. Harremoës, P., Johnson, O., Kontoyiannis, I.: Thinning and the law of small numbers. In: *IEEE International Symposium on Information Theory, 2007. ISIT 2007*, pp. 1491–1495. IEEE (2007)
3. Hillion, E., Johnson, O., Yu, Y.: Translations of probability measures on \mathbb{Z}
4. Léonard, C.: From the Schrödinger problem to the Monge–Kantorovich problem. *J. Funct. Anal.* (2011)
5. Lott, J., Villani, C.: Ricci curvature for metric-measure spaces via optimal transport. *Ann. Math.* **169**, 903–991
6. Ohta, S.: On the measure contraction property of metric measure spaces. *Commentarii Mathematici Helvetici* **82**(4), 805–828 (2007)
7. Rényi, A.: A characterization of poisson processes. *Magyar Tud. Akad. Mat. Kutató. Int. Közl* **1**, 519–527 (1956)
8. Sturm, K.T.: On the geometry of metric measure spaces I. *Acta Math.* **196**(1), 65–131 (2006)
9. Sturm, K.T.: On the geometry of metric measure spaces II. *Acta Math.* **196**(1), 133–177 (2006)
10. Villani, C.: *Topics in optimal transportation*, vol. 58. Amer Mathematical Society (2003)
11. Villani, C.: *Optimal transport: old and new*, vol. 338. Springer, Berlin (2009)
12. Yu, Y., Johnson, O.: Concavity of entropy under thinning. In: *IEEE International Symposium on Information Theory, 2009. ISIT 2009*, pp. 144–148. IEEE (2009)