



HAL
open science

Transport-entropy inequalities and curvature in discrete-space Markov chains

Ronen R Eldan, James R Lee, Joseph Lehec

► **To cite this version:**

Ronen R Eldan, James R Lee, Joseph Lehec. Transport-entropy inequalities and curvature in discrete-space Markov chains. *A journey through discrete mathematics*, Springer, pp.391-406, 2018. hal-01428953

HAL Id: hal-01428953

<https://hal.science/hal-01428953>

Submitted on 6 Jan 2017

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

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

Transport-entropy inequalities and curvature in discrete-space Markov chains

Ronen Eldan

James R. Lee

Joseph Lehec

Abstract

Let $G = (\Omega, E)$ be a graph and let d be the graph distance. Consider a discrete-time Markov chain $\{Z_t\}$ on Ω whose kernel p satisfies $p(x, y) > 0 \Rightarrow \{x, y\} \in E$ for every $x, y \in \Omega$. In words, transitions only occur between neighboring points of the graph. Suppose further that (Ω, p, d) has coarse Ricci curvature at least $1/\alpha$ in the sense of Ollivier: For all $x, y \in \Omega$, it holds that

$$W_1(Z_1 \mid \{Z_0 = x\}, Z_1 \mid \{Z_0 = y\}) \leq \left(1 - \frac{1}{\alpha}\right) d(x, y),$$

where W_1 denotes the Wasserstein 1-distance.

In this note, we derive a transport-entropy inequality: For any measure μ on Ω , it holds that

$$W_1(\mu, \pi) \leq \sqrt{\frac{2\alpha}{2 - 1/\alpha} D(\mu \parallel \pi)},$$

where π denotes the stationary measure of $\{Z_t\}$ and $D(\cdot \parallel \cdot)$ is the relative entropy.

Peres and Tetali have conjectured a stronger consequence of coarse Ricci curvature, that a modified log-Sobolev inequality (MLSI) should hold, in analogy with the setting of Markov diffusions. We discuss how our approach suggests a natural attack on the MLSI conjecture.

1 Introduction

In geometric analysis on manifolds, it is by now well-established that the Ricci curvature of the underlying manifold has profound consequences for functional inequalities and the rate of convergence of Markov semigroups toward equilibrium. One can consult, in particular the books [BGL14] and [Vil09]. Indeed, in the setting of *diffusions* (see [BGL14, §1.11]), there is an elegant theory around the Bakry-Emery *curvature-dimension* condition.

Roughly speaking, in the setting of diffusion on a continuous space, when there is an appropriate “integration by parts” formula (that connects the Dirichlet form to the Laplacian), a positive curvature condition implies powerful functional inequalities. Most pertinent to the present discussion, positive curvature yields transport-entropy and logarithmic Sobolev inequalities.

For discrete state spaces, the situation appears substantially more challenging. There are numerous attempts at generalizing lower bounds on the Ricci curvature to discrete metric measure spaces; we refer to the upcoming survey [CGM⁺16]. At a broad level, these approaches suffer from one of two drawbacks: Either the notion of “positive curvature” is difficult to verify for concrete spaces, or the “expected” functional analytic consequences do not follow readily.

In the present note, we consider the notion of *coarse Ricci curvature* due to Ollivier [Oll09]. It constitutes an approach of the latter type: There is a large body of finite-state Markov chains that have positive curvature in Ollivier’s sense, but for many of them we do not yet know if strong functional-analytic consequences hold. This study is made more fascinating by the straightforward connection between coarse Ricci curvature on graphs and the notion of *path coupling* arising in

the study of rapid mixing of Markov chains [BD97]. This is a powerful method to establish fast convergence to the stationary measure; see, for example, [LPW09, Ch. 14].

In particular, if there were an analogy to the diffusion setting that allowed coarse Ricci curvature lower bounds to yield logarithmic Sobolev inequalities (or variants thereof), it would even imply new mixing time bounds for well-studied chains arising from statistical physics, combinatorics, and theoretical computer science. A conjecture of Peres and Tetali asserts that a *modified log-Sobolev inequality* (MLSI) should always hold in this setting. Roughly speaking, this means that the underlying Markov chain has exponential convergence to equilibrium in the relative entropy distance.

Our aim is to give some preliminary results in this direction and to suggest a new approach to establishing MLSI. In particular, we prove a W_1 transport-entropy inequality. By results of Bobkov and Götze [BG99], this is equivalent to a sub-Gaussian concentration estimate for Lipschitz functions. Sammer has shown that such an inequality follows formally from MLSI [Sam05], thus one can see verification as evidence in support of the Peres-Tetali conjecture. Our result also addresses PROBLEM J in Ollivier’s survey [Oll10].

1.1 Coarse Ricci curvature and transport-entropy inequalities

Let Ω be a countable state space, and let $p : \Omega \times \Omega \rightarrow [0, 1]$ denote a transition kernel. For $x \in \Omega$, we will use the notation $p(x, \cdot)$ to denote the function $y \mapsto p(x, y)$. For a probability measure π on Ω and $f : \Omega \rightarrow \mathbb{R}_+$, we define the entropy of f by

$$\text{Ent}_\pi(f) = \mathbb{E}_\pi \left[f \log \left(\frac{f}{\mathbb{E}_\pi[f]} \right) \right].$$

We also equip Ω with a metric d . If μ and ν are two probability measures on Ω , we denote by $W_1(\mu, \nu)$ the transportation cost (or Wasserstein 1-distance) between μ and ν , with the cost function given by the distance d . Namely,

$$W_1(\mu, \nu) = \inf \{ \mathbb{E} [d(X, Y)] \}$$

where the infimum is taken on all couplings (X, Y) of (μ, ν) . Recall the Monge–Kantorovitch duality formula for W_1 (see, for instance, [Vil09, Case 5.16]):

$$W_1(\mu, \nu) = \sup \left\{ \int_\Omega f \, d\mu - \int_\Omega f \, d\nu \right\}, \tag{1}$$

where the supremum is taken over 1-Lipschitz functions f . We consider the following notion of curvature introduced by Ollivier [Oll09].

Definition 1.1. The *coarse Ricci curvature* of (Ω, p, d) is the largest $\kappa \in [-\infty, 1]$ such that the inequality

$$W_1(p(x, \cdot), p(y, \cdot)) \leq (1 - \kappa) d(x, y)$$

holds true for every $x, y \in \Omega$.

In the sequel we will be interested in positive Ricci curvature. Under this condition the map $\mu \mapsto \mu p$ is a contraction for W_1 . As a result, it has a unique fixed point and μp^n converges to this fixed point as $n \rightarrow \infty$. In other words the Markov kernel p has a unique stationary measure and is ergodic. The main purpose of this note is to show that positive curvature yields a transport-entropy inequality, or equivalently a Gaussian concentration inequality for the stationary measure.

Definition 1.2. We say that a probability measure μ on Ω satisfies the *Gaussian concentration property* with constant C if the inequality

$$\int_{\Omega} \exp(f) d\mu \leq \exp\left(\int_{\Omega} f d\mu + C \|f\|_{\text{Lip}}^2\right)$$

holds true for every Lipschitz function f .

Now we spell out the dual formulation of the Gaussian concentration property in terms of transport inequality. Recall first the definition of the relative entropy (or Kullback divergence): for two measures μ, ν on (Ω, \mathcal{B}) ,

$$D(\nu \parallel \mu) = \text{Ent}_{\mu}\left[\frac{d\nu}{d\mu}\right] = \int_{\Omega} \log\left(\frac{d\nu}{d\mu}\right) d\nu$$

if ν is absolutely continuous with respect to μ and $D(\nu \parallel \mu) = +\infty$ otherwise. As usual, if X and Y are random variables with laws ν and μ , we will take $D(X \parallel Y)$ to be synonymous with $D(\nu \parallel \mu)$.

Definition 1.3. We say that μ satisfies (T_1) with constant C if for every probability measure ν on Ω we have

$$W_1(\mu, \nu)^2 \leq C \cdot D(\nu \parallel \mu). \quad (T_1)$$

As observed by Bobkov and Götze [BG99], the inequality (T_1) and the Gaussian concentration property are equivalent.

Lemma 1.4. *A probability measure μ satisfies the Gaussian concentration property with constant C if and only if it satisfies (T_1) with constant $4C$.*

This is a relatively straightforward consequence of the Monge–Kantorovitch duality (1); we refer to [BG99] for details.

Theorem 1.5. *Assume that (Ω, p, d) has positive coarse Ricci curvature $1/\alpha$ and that the one-step transitions all satisfy (T_1) with the same constant C : Suppose that for every $x \in \Omega$ and for every probability measure ν we have*

$$W_1(\nu, p(x, \cdot))^2 \leq C \cdot D(\nu \parallel p(x, \cdot)). \quad (2)$$

Then the stationary measure π satisfies T_1 with constant $\frac{C\alpha}{2-1/\alpha}$.

Remark 1.6. Observe that Theorem 1.5 does not assume reversibility.

The hypothesis (2) might seem unnatural at first sight but it is automatically satisfied for the random walk on a graph when d is the graph distance. Indeed, recall Pinsker’s inequality: For every probability measures μ, ν we have

$$\text{TV}(\mu, \nu) \leq \sqrt{\frac{1}{2} D(\mu \parallel \nu)},$$

where TV denotes the total variation distance. This yields the following lemma.

Lemma 1.7. *Let μ be a probability measure on a metric space (M, d) and assume that the support of μ has finite diameter Δ . Then μ satisfies (T_1) with constant $\Delta^2/2$.*

Proof. Let ν be absolutely continuous with respect to μ . Then both μ and ν are supported on a set of diameter Δ . This implies that

$$W_1(\mu, \nu) \leq \Delta \cdot \text{TV}(\mu, \nu).$$

Combining this with Pinsker’s inequality we get $W_1(\mu, \nu)^2 \leq \frac{\Delta^2}{2} D(\nu \parallel \mu)$, which is the desired result. \square

Random walks on graphs. A particular case of special interest will be random walks on finite graphs. Let $G = (V, E)$ be a connected, undirected graph, possibly with self-loops. Given non-negative conductances $c : E \rightarrow \mathbb{R}_+$ on the edges, we recall the Markov chain $\{X_t\}$ defined by

$$\Pr[X_{t+1} = y \mid X_t = x] = \frac{c(\{x, y\})}{\sum_{z \in V} c(\{x, z\})}.$$

We refer to any such chain as *a random walk on the graph G* . If it holds that $c(\{x, x\}) \geq \frac{1}{2} \sum_{z \in V} c(\{x, z\})$ for all $x \in V$, we say that the corresponding random walk is *lazy*. We will equip G with its graph distance d .

In this setting, the transitions of the walk are supported on a set of diameter 2. So combining the preceding lemma with Theorem 1.5, one arrives at the following.

Corollary 1.8. *If a random walk on a graph has positive coarse Ricci curvature $\frac{1}{\alpha}$ (with respect to the graph distance), then the stationary measure π satisfies*

$$W_1(\mu, \pi)^2 \leq \frac{2\alpha}{2 - 1/\alpha} D(\mu \parallel \pi),$$

for every probability measure μ .

Remark 1.9. One should note that in this context we have

$$d(x, y) \leq W_1(p(x, \cdot), p(y, \cdot)) + 2, \quad \forall x, y \in \Omega,$$

just because after one step the walk is at distance 1 at most from its starting point. As a result, having coarse Ricci curvature $1/\alpha$ implies that the diameter Δ of the graph is at most 2α . So by the previous lemma, *every* measure on the graph satisfies T_1 with constant $2\alpha^2$. The point of Corollary 1.8 is that for the stationary measure π the constant is order α rather than α^2 .

We now present two proofs of Theorem 1.5. The first proof is rather short and based on the duality formula (1). The second argument provides an explicit coupling based on an entropy-minimal drift process. In Section 3, we discuss logarithmic Sobolev inequalities. In particular, we present a conjecture about the structure of the entropy-minimal drift that is equivalent to the Peres-Tetali MLSI conjecture.

After the first version of this note was released we were notified that Theorem 1.5 was proved by Djellout, Guillin and Wu in [DGW04, Proposition 2.10]. Note that this article actually precedes Ollivier's work. The proof given there corresponds to our first proof, by duality. Our second proof is more original but does share some similarities with the argument given by K. Marton in [Mar96, Proposition 1]. Also, after hearing about our work, Fathi and Shu [FS15] used their transport-information framework to provide yet another proof.

2 The W_1 transport-entropy inequality

We now present two proofs of Theorem 1.5. Recall the relevant data (Ω, p, d) . Define the process $\{B_t\}$ to be the discrete-time random walk on Ω corresponding to the transition kernel p . For $x \in \Omega$, we will use $B_t(x)$ to denote the random variable $B_t \mid \{B_0 = x\}$. For $t \geq 0$, we make the definition

$$P_t[f](x) = \mathbb{E}[f(B_t(x))].$$

2.1 Proof by duality

Let $f : \Omega \rightarrow \mathbb{R}$ be a Lipschitz function. Using the hypothesis (2) and Lemma 1.4 we get

$$P_1[\exp(f)](x) \leq \exp\left(P_1[f](x) + \frac{C}{4}\|f\|_{\text{Lip}}^2\right),$$

for all $x \in \Omega$. Applying this inequality repeatedly we obtain

$$P_n[\exp(f)](x) \leq \exp\left(P_n[f](x) + \frac{C}{4}\sum_{k=0}^{n-1}\|P_k f\|_{\text{Lip}}^2\right), \quad (3)$$

for every integer n and all $x \in \Omega$. Now we use the curvature hypothesis. Note that the Monge–Kantorovitch duality (1) yields easily

$$\begin{aligned} 1 - \frac{1}{\alpha} &= \sup_{x \neq y} \left\{ \frac{W_1(p(x, \cdot), p(y, \cdot))}{d(x, y)} \right\} \\ &= \sup_{x \neq y, g} \left\{ \frac{P_1[g](x) - P_1[g](y)}{\|g\|_{\text{Lip}} d(x, y)} \right\} \\ &= \sup_g \left\{ \frac{\|P_1[g]\|_{\text{Lip}}}{\|g\|_{\text{Lip}}} \right\}. \end{aligned}$$

Therefore $\|P_1[g]\|_{\text{Lip}} \leq (1 - 1/\alpha)\|g\|_{\text{Lip}}$ for every Lipschitz function g and thus

$$\|P_n[f]\|_{\text{Lip}} \leq (1 - 1/\alpha)^n \|f\|_{\text{Lip}},$$

for every integer n . Inequality (3) then yields

$$P_n[\exp(f)](x) \leq \exp\left(P_n[f](x) + \frac{C\alpha}{4(2 - 1/\alpha)}\|f\|_{\text{Lip}}^2\right).$$

Letting $n \rightarrow \infty$ yields

$$\int_{\Omega} \exp(f) d\pi \leq \exp\left(\int_{\Omega} f d\pi + \frac{C\alpha}{4(2 - 1/\alpha)}\|f\|_{\text{Lip}}^2\right).$$

The stationary measure π thus satisfies Gaussian concentration with constant $\frac{C\alpha}{4(2 - 1/\alpha)}$. Another application of the duality, Lemma 1.4, yields the desired outcome, proving Theorem 1.5.

2.2 An explicit coupling

As promised, we now present a second proof of Theorem 1.5 based on an explicit coupling. The proof does not rely on duality, and our hope is that the method presented will be useful for establishing MLSI; see Section 3.

The first step of the proof follows a similar idea to the one used in [Mar96, Proposition 1]. Given the random walk $\{B_t\}$ and another process $\{X_t\}$ (not necessarily Markovian), there is a natural coupling between the two processes that takes advantage of the curvature condition and gives a bound on the distance between the processes at time T in terms of the relative entropy. This step is summarized in the following result.

Proposition 2.1. *Assume that (Ω, p, d) satisfies the conditions of Theorem 1.5. Fix a time T and a point $x_0 \in M$. Let $\{B_0 = x_0, B_1, \dots, B_T\}$ be the corresponding discrete time random walk starting from x_0 and let $\{X_0 = x_0, X_1, \dots, X_T\}$ be an arbitrary random process on Ω starting from x_0 . Then, there exists a coupling between the processes (X_t) and (B_t) such that*

$$E[d(X_T, B_T)] \leq \sqrt{\frac{C\alpha}{2 - 1/\alpha} D(\{X_0, X_1, \dots, X_T\} \| \{B_0, B_1, \dots, B_T\})}.$$

In view of the above proposition, proving a transportation-entropy inequality for (Ω, p, d) is reduced to the following: given a measure ν on Ω , we are looking for a process $\{X_t\}$ which satisfies: (i) $X_T \sim \nu$ and (ii) the relative entropy between (X_0, \dots, X_T) and (B_0, \dots, B_T) is as small as possible.

To achieve the above, our key idea is the construction of a process X_t which is entropy minimal in the sense that it satisfies

$$X_T \sim \nu \text{ and } D(\{X_0, X_1, \dots, X_T\} \| \{B_0, B_1, \dots, B_T\}) = D(X_T \| B_T). \quad (4)$$

This process can be thought of as the Doob transform of the random walk with a given target law. In the setting of Brownian motion on \mathbb{R}^n equipped with the Gaussian measure, the corresponding process appears in work of Föllmer [Föll85, Föll86]. See [Leh13] for applications to functional inequalities, and the work of Léonard [Léo14] for a somewhat different perspective on the connection to optimal transportation.

2.2.1 Proof of Proposition 2.1: Construction of the coupling

Given $t \in \{1, \dots, T\}$ and $x_1, \dots, x_{t-1} \in M$, let $\nu(t, x_0, \dots, x_{t-1}, \cdot)$ be the conditional law of X_t given $X_0 = x_0, \dots, X_{t-1} = x_{t-1}$. Now we construct the coupling of X and B as follows. Set $X_0 = B_0 = x_0$ and given $(X_1, B_1), \dots, (X_{t-1}, B_{t-1})$ set (X_t, B_t) to be a coupling of $\nu(t, X_0, \dots, X_{t-1}, \cdot)$ and $p(B_{t-1}, \cdot)$ which is optimal for W_1 . Then by construction the marginals of this process coincide with the original processes $\{X_t\}$ and $\{B_t\}$.

The next lemma follows from the coarse Ricci curvature property and the definition of our coupling.

Lemma 2.2. *For every $t \in \{1, \dots, T\}$,*

$$\mathbb{E}_{t-1} [d(X_t, B_t)] \leq \sqrt{C \cdot D(\nu(t, X_0, \dots, X_{t-1}, \cdot) \| p(X_{t-1}, \cdot))} + \left(1 - \frac{1}{\alpha}\right) d(X_{t-1}, B_{t-1})$$

where $\mathbb{E}_{t-1}[\cdot]$ stands for the conditional expectation given $(X_0, B_0), \dots, (X_{t-1}, B_{t-1})$.

Proof. By definition of the coupling, the triangle inequality for W_1 , the one-step transport inequality (2) and the curvature condition

$$\begin{aligned} \mathbb{E}_{t-1} [d(X_t, B_t)] &= W_1(\nu(t, X_0, \dots, X_{t-1}, \cdot), p(B_{t-1}, \cdot)) \\ &\leq W_1(\nu(t, X_0, \dots, X_{t-1}, \cdot), p(X_{t-1}, \cdot)) + W_1(p(X_{t-1}, \cdot), p(B_{t-1}, \cdot)) \\ &\leq \sqrt{C \cdot D(\nu(t, X_0, \dots, X_{t-1}, \cdot) \| p(X_{t-1}, \cdot))} + \left(1 - \frac{1}{\alpha}\right) d(X_{t-1}, B_{t-1}). \quad \square \end{aligned}$$

Remark that the chain rule for relative entropy asserts that

$$\sum_{t=1}^T \mathbb{E}[D(\nu(t, X_0, \dots, X_{t-1}, \cdot) \| p(X_{t-1}, \cdot))] = D(\{X_0, X_1, \dots, X_T\} \| \{B_0, B_1, \dots, B_T\}). \quad (5)$$

Using the preceding lemma inductively and then Cauchy-Schwarz yields

$$\begin{aligned} \mathbb{E}[d(X_T, B_T)] &\leq \sum_{t=1}^T \left(1 - \frac{1}{\alpha}\right)^{T-t} \mathbb{E} \left[\sqrt{C \cdot D(\nu(t, X_0, \dots, X_{t-1}, \cdot) \| p(X_{t-1}, \cdot))} \right] \\ &\leq \sqrt{\sum_{t=1}^T \left(1 - \frac{1}{\alpha}\right)^{2(T-t)}} \sqrt{\sum_{t=1}^T C \cdot \mathbb{E}[D(\nu(t, X_0, \dots, X_{t-1}, \cdot) \| p(X_{t-1}, \cdot))]} \\ &\stackrel{(5)}{\leq} \sqrt{\frac{\alpha}{2 - 1/\alpha}} \sqrt{C \cdot D(\{X_0, X_1, \dots, X_T\} \| \{B_0, B_1, \dots, B_T\})}, \end{aligned}$$

completing the proof of Proposition 2.1.

2.2.2 The entropy-optimal drift process

Our goal in this section is to construct a process $x_0 = X_0, X_1, \dots, X_T$ satisfying equation (4). Suppose that we are given a measure ν on Ω along with an initial point $x_0 \in \Omega$ and a time $T \geq 1$. We define the *Föllmer drift process* associated to (ν, x_0, T) as the stochastic process $\{X_t\}_{t=0}^T$ defined as follows.

Let μ_T be the law of $B_T(x_0)$ and denote by f the density of ν with respect to μ_T . Note that f is well-defined as long as the support of μ_T is Ω . Now let $\{X_t\}_{t=0}^T$ be the non homogeneous Markov chain on Ω whose transition probabilities at time t are given by

$$q_t(x, y) := \mathbb{P}(X_t = y \mid X_{t-1} = x) = \frac{P_{T-t}f(y)}{P_{T-t+1}f(x)} p(x, y). \quad (6)$$

We will take care in what follows to ensure the denominator does not vanish. Note that (q_t) is indeed a transition matrix as

$$\sum_{y \in \Omega} P_{T-t}f(y) p(x, y) = P_{T-t+1}f(x). \quad (7)$$

We now state a key property of the drift.

Lemma 2.3. *If $p^T(x_0, x) > 0$ for all $x \in \text{supp}(\mu)$, then $\{X_t\}$ is well-defined. Furthermore, for every $x_1, \dots, x_T \in \Omega$ we have*

$$\mathbb{P}((X_1, \dots, X_T) = (x_1, \dots, x_T)) = \mathbb{P}((B_1, \dots, B_T) = (x_1, \dots, x_T)) f(x_T). \quad (8)$$

In particular X_T has law $d\nu = f d\mu_T$.

Proof. By definition of the process (X_t) we have

$$\begin{aligned} \mathbb{P}((X_1, \dots, X_T) = (x_1, \dots, x_T)) &= \prod_{t=1}^T \mathbb{P}(X_t = x_t \mid (X_1, \dots, X_{t-1}) = (x_1, \dots, x_{t-1})) \\ &= \prod_{t=1}^T \frac{P_{T-t}f(x_t)}{P_{T-t+1}f(x_{t-1})} p(x_{t-1}, x_t) \\ &= \frac{f(x_T)}{P_T f(x_0)} \left(\prod_{t=1}^T p(x_{t-1}, x_t) \right), \end{aligned}$$

which is the result. \square

In words, the preceding lemma asserts that the law of the process $\{X_t\}$ has density $f(x_T)$ with respect to the law of the process $\{B_t\}$. As a result we have in particular

$$D(\{X_0, X_1, \dots, X_T\} \parallel \{B_0, B_1, \dots, B_T\}) = \mathbb{E}[\log f(X_T)] = D(\nu \parallel \mu_T),$$

since X_T has law μ . Note that for any other process $\{Y_t\}$ such that $Y_0 = x_0$ and Y_T has law ν , one always has the inequality

$$D(\{Y_0, Y_1, \dots, Y_T\} \parallel \{B_0, B_1, \dots, B_T\}) \geq D(Y_T \parallel B_T) = D(\nu \parallel \mu_T). \quad (9)$$

Besides $\{X_t\}$ is the unique random process for which this inequality is tight. Uniqueness follows from strict convexity of the relative entropy.

We summarize this section in the following lemma.

Lemma 2.4. *Let (Ω, p) be a Markov chain. Fix $x_0 \in \Omega$ and let $x_0 = B_0, B_1, \dots$ be the associated random walk. Let ν be a measure on Ω and let $T > 0$ be such that for any $y \in \Omega$ one has that $\mathbb{P}(B_T = y) > 0$. Then there exists a process $x_0 = X_0, X_1, \dots, X_T$ such that:*

- $\{X_0, \dots, X_T\}$ is a (time inhomogeneous) Markov chain.
- X_T is distributed with the law ν .
- The process satisfies equation (4), namely

$$D(X_T \| B_T) = D(\{X_0, X_1, \dots, X_T\} \| \{B_0, B_1, \dots, B_T\}).$$

2.3 Finishing up the proof

Fix an arbitrary $x_0 \in \Omega$ and consider some $T \geq \text{diam}(\Omega, d)$. Let $\{X_t\}$ be the Föllmer drift process associated to the initial data (ν, x_0, T) . Then combining Proposition 2.1 and equation (4), we have

$$W_1(\nu, \mu_T) = \mathbb{E}[d(X_T, B_T)] \leq \sqrt{\frac{C\alpha}{2 - 1/\alpha} D(\nu \| \mu_T)}. \quad (10)$$

Now let $T \rightarrow \infty$ so that $\mu_T \rightarrow \pi$, yielding the desired claim.

3 The Peres-Tetali conjecture and log-Sobolev inequalities

Recall that $p : \Omega \times \Omega \rightarrow [0, 1]$ is a transition kernel on the finite state space Ω with a unique stationary measure π . Let $L^2(\Omega, \pi)$ denote the space of real-valued functions $f : \Omega \rightarrow \mathbb{R}$ equipped with the inner product $\langle f, g \rangle = \mathbb{E}_\pi[fg]$. From now on we assume that the measure π is reversible, which amounts to saying that the operator $f \mapsto pf$ is self-adjoint in $L^2(\Omega, \pi)$.

We define the associated Dirichlet form

$$\mathcal{E}(f, g) = \langle f, (p - I)g \rangle = \frac{1}{2} \sum_{x, y \in \Omega} \pi(x)p(x, y)(f(x) - f(y))(g(x) - g(y)).$$

Recall the definition of the entropy of a function $f : \Omega \rightarrow \mathbb{R}_+$:

$$\text{Ent}_\pi(f) = \mathbb{E}_\pi \left[f \log \left(\frac{f}{\mathbb{E}_\pi f} \right) \right].$$

Now define the quantities

$$\rho = \inf_{f: \Omega \rightarrow \mathbb{R}_+} \frac{\mathcal{E}(\sqrt{f}, \sqrt{f})}{\text{Ent}_\pi(f)}$$

$$\rho_0 = \inf_{f: \Omega \rightarrow \mathbb{R}_+} \frac{\mathcal{E}(f, \log f)}{\text{Ent}_\pi(f)}.$$

These numbers are called, respectively, the *log-Sobolev* and *modified log-Sobolev* constants of the chain (Ω, p) . We refer to [MT06] for a detailed discussion of such inequalities on discrete-space Markov chains and their relation to mixing times.

One can understand both numbers as measuring the rate of convergence to equilibrium in appropriate senses. The modified log-Sobolev constant, in particular, can be equivalently characterized as the largest value ρ_0 such that

$$\text{Ent}_\pi(H_t f) \leq e^{-\rho_0 t} \text{Ent}_\pi(f) \quad (11)$$

for all $f : \Omega \rightarrow \mathbb{R}_+$ and $t > 0$ (see [MT06, Prop. 1.7]). Here, $H_t : L^2(\Omega, \pi) \rightarrow L^2(\Omega, \pi)$ is the heat-flow operator associated to the *continuous-time* random walk, i.e., $H_t = e^{-t(I-P)}$, where P is the operator defined by $Pf(x) = \sum_{y \in \Omega} p(x, y)f(y)$.

The log-Sobolev constant ρ controls the hypercontractivity of the semigroup (H_t) , which in turn yields a stronger notion of convergence to equilibrium; again see [MT06] for a precise statement. Interestingly, in the setting of diffusions, there is no essential distinction between the two notions; one should consider the following calculation only in a formal sense:

$$\text{“ } \mathcal{E}(f, \log f) = \int \nabla f \nabla \log f = \int \frac{|\nabla f|^2}{f} = 4 \int |\nabla \sqrt{f}|^2 = 4 \mathcal{E}(\sqrt{f}, \sqrt{f}) \text{.”}$$

However, in the discrete-space setting, the tools of differential calculus are not present. Indeed, one has the bound $\rho \leq 2\rho_0$ [MT06, Prop 1.10], but there is no uniform bound on ρ_0 in terms of ρ .

3.1 MLSI and curvature

We are now in position to state an important conjecture linking curvature and the modified log-Sobolev constant; it asserts that on spaces with positive coarse Ricci curvature, the random walk should converge to equilibrium exponentially fast in the relative entropy distance.

Conjecture 3.1 (Peres-Tetali, unpublished). *Suppose (Ω, p) corresponds to lazy random walk on a finite graph and d is the graph distance. If (Ω, p, d) has coarse Ricci curvature $\kappa > 0$, then the modified log-Sobolev constant satisfies*

$$\rho_0 \geq C\kappa. \tag{12}$$

where $C > 0$ is a universal constant.

A primary reason for our interest in Corollary 1.8 is that, by results of Sammer [Sam05], Conjecture 3.1 implies Corollary 1.8. We suspect that a stronger conclusion should hold in many cases; under stronger assumptions, it should be that one can obtain a lower bound on the (non-modified) log-Sobolev constant $\rho \geq C\kappa$. See, for instance, the beautiful approach of Marton [Mar15] that establishes a log-Sobolev inequality for product spaces assuming somewhat strong contraction properties of the Gibbs sampler.

However, we recall that this cannot hold under just the assumptions of Conjecture 3.1. Indeed, if $G = (V, E)$ is the complete graph on n vertices, it is easy to see that the coarse Ricci curvature κ of the lazy random walk is $1/2$. On the other hand, one can check that the log-Sobolev constant ρ decays asymptotically like $\frac{1}{\log n}$ (use the test function $f = \delta_x$ for some fixed $x \in V$).

3.2 An entropic interpolation formulation of MLSI

We now suggest an approach to Conjecture 3.1 using an entropy-optimal drift process. While we chose to work with discrete-time chains in Section 2.2.2, working in continuous-time will allow us more precision in exploring Conjecture 3.1. We will use the notation introduced at the beginning of this section.

A continuous-time drift process. Suppose we have some initial data (f, x_0, T) where $x_0 \in \Omega$ and $f : \Omega \rightarrow \mathbb{R}_+$ satisfies $\mathbb{E}_\pi[f] = 1$. Let $\{B_t : t \in [0, \infty)\}$ denote the continuous-time random walk with jump rates p on the discrete state space Ω starting from x_0 . We let μ_T be the law of B_T and let ν be the probability measure defined by

$$d\nu = \frac{f}{H_T f(x_0)} d\mu_T,$$

where (H_t) is the semigroup associated to the jump rates $p(x, y)$. Note that ν is indeed a probability measure as $\int f d\mu_T = H_T f(x_0)$ by definition of μ_T .

We now define the continuous-time Föllmer drift process associated to the data (x_0, T, f) as the (time inhomogeneous) Markov chain $\{X_t, t \leq T\}$ starting from x_0 and having transition rates at time t given by

$$q_t(x, y) = p(x, y) \frac{H_{T-t}f(y)}{H_{T-t}f(x)}. \quad (13)$$

Informally this means that the conditional probability that the process $\{X_t\}$ jumps from x to y between time t and $t+dt$ given the past is $q_t(x, y)dt$. This should be thought as the continuous-time analogue of the discrete Föllmer process defined by (6). We claim that again the law of the process $\{X_t, t \leq T\}$ has density $f(x_T)/H_Tf(x_0)$ with respect to the law of $\{B_t, t \leq T\}$. Let us give a brief justification of this claim. Define a new probability measure \mathbb{Q} by setting

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \frac{f(B_T)}{H_Tf(x_0)}. \quad (14)$$

We want to prove that the law of B under \mathbb{Q} coincides with the law of X under \mathbb{P} . Let (\mathcal{F}_t) be the natural filtration of the process (B_t) , let $t \in [0, T)$ and let $y \in M$. We then have the following computation:

$$\begin{aligned} \mathbb{Q}(B_{t+\Delta t} = y \mid \mathcal{F}_t) &= \frac{\mathbb{E}^{\mathbb{P}}[f(B_T) \mathbf{1}_{\{B_{t+\Delta t}=y\}} \mid \mathcal{F}_t]}{\mathbb{E}^{\mathbb{P}}[f(B_T) \mid \mathcal{F}_t]} + o(\Delta t) \\ &= \frac{H_{T-t}f(y)}{H_{T-t}f(B_t)} \mathbb{P}(B_{t+\Delta t} = y \mid \mathcal{F}_t) + o(\Delta t) \\ &= \frac{H_{T-t}f(y)}{H_{T-t}f(B_t)} p(B_t, y) \Delta t + o(\Delta t). \end{aligned}$$

This shows that under \mathbb{Q} , the process $\{B_t, t \leq T\}$ is Markovian (non homogeneous) with jump rates at time t given by (13). Hence the claim.

This implies in particular that X_T has law ν . This also yields the following formula for the relative entropy of $\{X_t\}$:

$$D(\{X_t, t \leq T\} \parallel \{B_t, t \leq T\}) = \mathbb{E} \left[\log \frac{f(X_T)}{H_Tf(x_0)} \right] = D(\nu \parallel \mu_T). \quad (15)$$

The process $\{X_t\}$ starts from x_0 and has law ν at time T . Because X_T has law ν and B_T has law μ_T , the two processes must evolve differently. One can think of the process $\{X_t\}$ as “spending information” in order to achieve the discrepancy between X_T and B_T . The amount of information spent must at least account for the difference in laws at the endpoint, i.e.,

$$D(\{X_t, t \leq T\} \parallel \{B_t, t \leq T\}) \geq D(X_T \parallel B_T).$$

As pointed out in Section 2.2.2, the content of (15) is that $\{X_t\}$ spends exactly this minimum amount.

For $0 \leq s \leq s'$, we use the notations $B_{[s, s']} = \{B_t : t \in [s, s']\}$ and $X_{[s, s']} = \{X_t : t \in [s, s']\}$ for the corresponding trajectories. From the definition of \mathbb{Q} we easily get

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_t} = \mathbb{E} \left[\frac{f(B_T)}{H_Tf(x_0)} \mid \mathcal{F}_t \right] = \frac{H_{T-t}f(B_t)}{H_{T-t}f(x_0)}. \quad (16)$$

As a result

$$D(X_{[0, t]} \parallel B_{[0, t]}) = \mathbb{E} \left[\log \frac{H_{T-t}f(X_t)}{H_{T-t}f(x_0)} \right] \quad (17)$$

for all $t \leq T$. Let us now define the rate of information spent at time t :

$$I_t = \frac{d}{dt} D(X_{[0,t]} \| B_{[0,t]}) .$$

Intuitively, the entropy-optimal process $\{X_t\}$ will spend progressively more information as t approaches T . Information spent earlier in the process is less valuable (as the future is still uncertain). Let us observe that a formal version of this statement for random walks on finite graphs is equivalent to Conjecture 3.1.

Conjecture 3.2. *Suppose (Ω, p) corresponds to a lazy random walk on a finite graph and d is the graph distance, and that (Ω, p, d) has coarse Ricci curvature $1/\alpha$. Given $f : \Omega \rightarrow \mathbb{R}_+$ with $\mathbb{E}_\pi[f] = 1$ and $x_0 \in \Omega$, for all sufficiently large times T , it holds that if $\{X_t : t \in [0, T]\}$ is the associated continuous-time Föllmer drift process with initial data (f, x_0, T) , then*

$$D(X_T \| B_T) \leq C\alpha I_T , \tag{18}$$

where $C > 0$ is a universal constant.

As $T \rightarrow \infty$, we have $P_T f(x_0) \rightarrow \mathbb{E}_\pi f = 1$ and thus

$$D(X_T \| B_T) \rightarrow \text{Ent}_\pi(f)$$

Moreover, we claim that $I_T \rightarrow \mathcal{E}(f, \log f)$ as $T \rightarrow \infty$. Together, these show that Conjectures 3.1 and 3.2 are equivalent.

To verify the latter claim, note that from (17) and (16) we have

$$\begin{aligned} D(X_{[0,t]} \| B_{[0,t]}) &= \mathbb{E} \left[\log \left(\frac{H_{T-t} f(X_t)}{H_T f(x_0)} \right) \right] \\ &= \mathbb{E} \left[\log \left(\frac{H_{T-t} f(B_t)}{H_T f(x_0)} \right) \frac{H_{T-t} f(B_t)}{H_T f(x_0)} \right] \\ &= \frac{1}{H_T f(x_0)} H_t (H_{T-t} f \log H_{T-t} f)(x_0) - \log H_T f(x_0). \end{aligned}$$

Differentiating at $t = T$ yields

$$I_T = \frac{1}{H_T f(x_0)} H_T (\Delta(f \log f) - (\Delta f)(\log f + 1))(x_0).$$

where $\Delta = I - p$ denotes the generator of the semigroup (H_t) . Recall that $\delta_{x_0} H_T$ converges weakly to π , and that by stationarity $\mathbb{E}_\pi \Delta g = 0$ for every function g . Thus

$$\lim_{T \rightarrow \infty} I_T = -\mathbb{E}_\pi[(\Delta f) \log f].$$

The latter equals $\mathcal{E}(f, \log f)$ by reversibility, hence the claim.

References

- [BD97] Russ Bubley and Martin E. Dyer. Path coupling: A technique for proving rapid mixing in markov chains. In *38th Annual Symposium on Foundations of Computer Science, FOCS '97, Miami Beach, Florida, USA, October 19-22, 1997*, pages 223–231, 1997.
- [BG99] S. G. Bobkov and F. Götze. Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.*, 163(1):1–28, 1999.

- [BGL14] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*, volume 348 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Cham, 2014.
- [CGM⁺16] K. Costello, N. Gozlan, J. Melbourne, W. Perkins, C. Roberto, Y. Samson, Y. Shu, and P. Tetali. Discrete Ricci curvature and functional inequalities on graphs. In preparation, 2016.
- [DGW04] H. Djellout, A. Guillin, and L. Wu. Transportation cost-information inequalities and applications to random dynamical systems and diffusions. *Ann. Probab.*, 32(3B):2702–2732, 2004.
- [Föll85] H. Föllmer. An entropy approach to the time reversal of diffusion processes. In *Stochastic differential systems (Marseille-Luminy, 1984)*, volume 69 of *Lecture Notes in Control and Inform. Sci.*, pages 156–163. Springer, Berlin, 1985.
- [Föll86] H. Föllmer. Time reversal on Wiener space. In *Stochastic processes—mathematics and physics (Bielefeld, 1984)*, volume 1158 of *Lecture Notes in Math.*, pages 119–129. Springer, Berlin, 1986.
- [FS15] M. Fathi and Y. Shu. Curvature and transport inequalities for markov chains in discrete spaces. *Preprint: arXiv: 1509.07160*, 2015.
- [Leh13] Joseph Lehec. Representation formula for the entropy and functional inequalities. *Ann. Inst. Henri Poincaré Probab. Stat.*, 49(3):885–899, 2013.
- [Léo14] Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, 34(4):1533–1574, 2014.
- [LPW09] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2009. With a chapter by James G. Propp and David B. Wilson.
- [Mar96] K. Marton. Bounding \bar{d} -distance by informational divergence: a method to prove measure concentration. *Ann. Probab.*, 24(2):857–866, 1996.
- [Mar15] K. Marton. Logarithmic Sobolev inequalities in discrete product spaces: A proof by a transportation cost distance. *Preprint: arXiv: 1507.02803*, 2015.
- [MT06] Ravi Montenegro and Prasad Tetali. Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci.*, 1(3):x+121, 2006.
- [Oll09] Yann Ollivier. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.*, 256(3):810–864, 2009.
- [Oll10] Yann Ollivier. A survey of Ricci curvature for metric spaces and Markov chains. In *Probabilistic approach to geometry*, volume 57 of *Adv. Stud. Pure Math.*, pages 343–381. Math. Soc. Japan, Tokyo, 2010.
- [Sam05] Marcus D. Sammer. *Aspects of mass transportation in discrete concentration inequalities*. ProQuest LLC, Ann Arbor, MI, 2005. Thesis (Ph.D.)—Georgia Institute of Technology.
- [Vil09] Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 2009. Old and new.