DISCRETE RICCI CURVATURE BOUNDS FOR BERNOULLI–LAPLACE AND RANDOM TRANSPOSITION MODELS

MATTHIAS ERBAR, JAN MAAS, PRASAD TETALI

Abstract. We calculate a Ricci curvature lower bound for some classical examples of random walks, namely, a chain on a slice of the n-dimensional discrete cube (the so-called Bernoulli–Laplace model) and the random transposition shuffle of the symmetric group of permutations on n letters.

Contents

1. Introduction

Many analytic and probabilistic properties of diffusion processes can be derived from geometric properties of the underlying space. In particular, a positive lower bound on the *Ricci curvature* on a Riemannian manifold has significant consequences for the associated heat semigroup/Brownian motion. In fact, such a bound implies a logarithmic Sobolev inequality, a Poincaré inequality, and a Brunn–Minkowski inequality, as well as several geometric inequalities.

Because of this wide range of implications, major research activity has been devoted to developing a notion of Ricci curvature (lower boundedness) that applies to non-smooth settings. Several approaches have been developed. Bakry–Eméry [2] introduced an approach based on algebraic properties of diffusion operators (the so-called Γ_2-calculus). Later, an approach based on optimal transport has been developed by Lott, Sturm and Villani [24, 17], and subsequently refined by Ambrosio, Gigli and Savaré [1]. In recent years, the equivalence of

Date: October 1, 2014.

P.T. gratefully acknowledges support by the NSF grants DMS-1101447 and DMS-1407657. E.M and J.M gratefully acknowledge support by the German Research Foundation through the Collaborative Research Center 1060 *The Mathematics of Emergent Effects* and the Hausdorff Center for Mathematics. This material is based upon work supported by the National Science Foundation under Grant No. 0932078 000, while E.M and J.M were in residence at MSRI, in the fall of 2013. All three authors thank the Mathematical Sciences Research Institute (MSRI) and the Simons Institute for the Theory of Computing, Berkeley, CA, for the hospitality and the conducive atmosphere of these institutes which facilitated this research collaboration.

the algebraic approach and the optimal transport approach has been proved, and a complete picture is emerging.

However, since the theory does not apply to discrete settings, several discrete notions of Ricci curvature have been introduced. In particular, the notion of *coarse Ricci curvature* was developed in considerable detail by Ollivier [21], although the basics were implicit in Dobrushin's work and others' since (see, e.g., the discussion in [22]), besides the notion being made explicit in the Ph.D. thesis of Sammer [23]. This notion is based on contraction properties of a Markov kernel in the (Kantorovich) W_1-metric. In this paper we focus on a different notion of Ricci curvature, which was proposed in [18] and systematically studied in [9].

1.1. **A discrete notion of Ricci curvature.** Let L be the generator of a continuous time Markov chain on a finite set $\mathcal{X}$, thus for functions $\psi : \mathcal{X} \to \mathbb{R}$, the operator L is of the form $L\psi(x) = \sum_{y\in\mathcal{X}} Q(x,y)(\psi(y) - \psi(x))$ where $Q(x,y) \geq 0$ for all $x, y \in \mathcal{X}$ with $x \neq y$, and $Q(x,x) = 0$ for all $x \in \mathcal{X}$. We shall assume that there exists a reversible probability measure π on $\mathcal{X}$, which means that $\pi(x)Q(x,y) = \pi(y)Q(y,x)$ for all x, y. We let $\mathscr{P}(\mathcal{X}) = \{\rho \in \mathbb{R}_+^{\mathcal{X}} : \sum_x \rho(x)\pi(x) = 1\}$ be the space of probability densities on $\mathcal{X}$ and denote by $\mathcal{H}(\rho) = \sum_x \rho(x) \log \rho(x)\pi(x)$ the relative entropy of $\rho \in \mathscr{P}(\mathcal{X})$.

In [18] a metric $\mathcal{W}$ on the space of probability measures has been constructed with the property that the heat flow is the gradient flow of the relative entropy. In this sense, $\mathcal{W}$ may be regarded as a natural analogue of the 2-Wasserstein metric induced by the Markov triple $(\mathcal{X}, Q, \pi)$. We refer to Section 2 for the precise definition of $\mathcal{W}$.

We say that $(\mathcal{X}, Q, \pi)$ has *Ricci curvature bounded from below by* $\kappa \in \mathbb{R}$ if the relative entropy $\mathcal{H}$ is κ-geodesically convex along $\mathcal{W}$-geodesics. More explicitly, for any constant speed geodesic $\{\rho_t\}_{t\in[0,1]}$ in $(\mathscr{P}(\mathcal{X}), \mathcal{W})$, we require that

$$\mathcal{H}(\rho_t) \leq (1-t)\mathcal{H}(\rho_0) + t\mathcal{H}(\rho_1) - \frac{\kappa}{2} t(1-t)\mathcal{W}(\rho_0, \rho_1)^2 .$$

In this case, we write

$$\mathrm{Ric}(\mathcal{X}, Q, \pi) \geq \kappa .$$

This notion of Ricci curvature is a direct analogue of the notion introduced by Lott, Sturm, and Villani in the setting of geodesic metric measure spaces.

It has been shown in [9] that this notion of Ricci curvature has significant consequences, such as an HWI-inequality à la Otto–Villani, a modified logarithmic Sobolev inequality (MLSI) and a Poincaré (or spectral gap) inequality. The MLSI (with constant $\alpha > 0$) asserts that

$$\mathcal{H}(\rho) \leq \alpha^{-1}\mathcal{E}(\rho, \log \rho) ,$$

for all $\rho \in \mathscr{P}(\mathcal{X})$, where $\mathcal{E}$ is the associated Dirichlet form defined by

$$\mathcal{E}(f,g) \ = \ -\langle f, Lg\rangle_{L^2(\mathcal{X},\pi)} \ = \ \frac{1}{2}\sum_{x,y\in\mathcal{X}} \big(f(y) - f(x)\big)\big(g(y) - g(x)\big)Q(x,y)\pi(x) .$$

The MLSI is equivalent to the exponential convergence estimate $\mathcal{H}(e^{tL}\rho) \leq e^{-\alpha t}\mathcal{H}(\rho)$. The Poincaré inequality asserts that

$$\|\psi\|^2_{L^2(\mathcal{X},\pi)} \leq \lambda^{-1}\mathcal{E}(\psi,\psi) ,$$

for all functions $\psi : \mathcal{X} \to \mathbb{R}$ with $\sum_{x\in\mathcal{X}} \psi(x)\pi(x) = 0$. It is equivalent to the exponential convergence estimate $\|e^{tL}\psi\|_{L^2(\mathcal{X},\pi)} \leq e^{-\lambda t}\|\psi\|_{L^2(\mathcal{X},\pi)}$. From now on, we will denote the optimal constants in the inequalities by κ, α and λ respectively. It is well known that $\lambda \geq \alpha/2$. Moreover, it has been proved in [9] that $\alpha/2 \geq \kappa$.

In view of these consequences it is desirable to obtain sharp Ricci curvature bounds in concrete discrete examples. So far, very little is known in this direction. Two types of results have been obtained:

- Mielke [20] obtained Ricci curvature bounds for one-dimensional birth-death chains. He applies his bounds to approximations of Fokker–Planck equations with κ-convex potential and shows that the curvature of the discrete approximations converge to κ. The proof relies on diagonal dominance of the Hessian matrix, and seems to be restricted to 1-dimensional situations.
- Erbar–Maas [9] obtained a tensorisation result for Ricci curvature: if $\mathrm{Ric}(\mathcal{X}_i, Q_i, \pi_i) \geq \kappa_i$ for $i = 1, 2$, then the associated product chain on the product space $\mathcal{X}_1 \times \mathcal{X}_2$ has Ricci curvature bounded from below by $\min\{\kappa_1, \kappa_2\}$.

Apart from the elementary example of the complete graph, no results are available beyond the 1-dimensional or the product setting. This paper provides the first results in this direction.

In a different direction, Gozlan et al [15] constructed an interpolation on the space of probability measures and derived a displacement convexity of entropy inequality with respect to the classical W_1-metric on the complete graph and products of complete graphs, in particular, the n-dimensional discrete cube; the results thus obtained are consistent with the bounds on the curvature in the sense [9] as well as the coarse Ricci curvature.

1.2. **The Bernoulli-Laplace model.** The Bernoulli-Laplace model is the simple exclusion process on the complete graph and can be described as follows. Consider k indistinguishable particles distributed over n sites labeled by $[n] = \{1, \dots, n\}$, where $1 \leq k < n$. Each site contains at most one particle. The state space of the system is the set $\Omega(n,k) = \{x \in \{0,1\}^n : x_1 + \cdots + x_n = k\}$ (or equivalently, the set of all subsets of $[n]$ of size k).

The Bernoulli-Laplace model is the continuous time Markov chain described as follows: after random waiting times (independent exponentially distributed with rate $\frac{1}{k(n-k)}$), one particle is selected uniformly at random, and jumps to a free site, selected uniformly at random. The transition rates are thus given by

$$Q_{\mathrm{BL}}(x,y) = \begin{cases} \frac{1}{k(n-k)} \ , & \text{if } \|x-y\|_{\ell^1} = 2 \ , \\ 0 \ , & \text{otherwise} \ . \end{cases}$$

The uniform probability measure on $\Omega(n,k)$, given by $\pi_{\mathrm{BL}}(x) = \binom{n}{k}^{-1}$ for all x, is reversible for Q_{BL}. Note that the Bernoulli-Laplace model may be seen as the simple random walk on $\Omega(n,k)$ endowed with the Hamming distance $d(x,y) = \frac{1}{2}\|x-y\|_{\ell^1}$.

We prove the following result:

Theorem 1.1 (Ricci bound for the Bernoulli-Laplace model)**.** *Let $n > 1$ and $1 \leq k \leq n-1$. The Ricci curvature of the Bernoulli-Laplace model $(\Omega(n,k), Q_{\mathrm{BL}}, \pi_{\mathrm{BL}})$ is bounded from below by $\frac{n+2}{2k(n-k)}$.*

The mixing time for the Bernoulli-Laplace model has been studied by Diaconis and Shashahani [8], who showed in particular that the spectral gap equals $\frac{n}{k(n-k)}$. Their analysis is based on lifting the model to the symmetric group and using representation theory in this setting. Lee and Yau [16] obtained a sharp logarithmic Sobolev inequality, improving earlier work by

Diaconis and Saloff-Coste [6]. In three independent works Gao–Quastel [11], Goel [13], and Bobkov–Tetali [3] proved the following (lower) bound on the MLSI constant:

$$\frac{n}{2k(n-k)} \;\leq\; \alpha \;\leq\; \frac{2n}{k(n-k)}\;,$$

where the upper bound comes from the fact that $\alpha \leq 2\lambda$. Since $\alpha \geq 2\kappa$ by [9, Thm. 7.4], our Theorem 1.1 implies that

$$\frac{n+2}{k(n-k)} \;\leq\; \alpha \;\leq\; \frac{2n}{k(n-k)}\;,$$

which improves the lower bound above roughly by a factor 2. Such an improvement on the MLSI constant for the Bernoulli-Laplace model has previously been obtained by Caputo et al. [4].

1.3. **The random transposition model.** Let $n \geq 1$, and let $\mathcal{S}_n$ be the group of all permutations of $[n]$. We define a graph structure on $\mathcal{S}_n$ by connecting two permutations $\sigma_1, \sigma_2 \in \mathcal{S}_n$ if $\sigma_2 = \tau \circ \sigma_1$, for some transposition τ. (Recall that a transposition is a permutation that interchanges precisely two elements). In this case we write $\sigma_1 \sim \sigma_2$. Simple random walk is then defined by

$$Q_{\mathrm{RT}}(\sigma_1, \sigma_2) = \begin{cases} \frac{2}{n(n-1)}\;, & \text{if } \sigma_1 \sim \sigma_2\;, \\ 0\;, & \text{otherwise}\;. \end{cases}$$

The uniform measure π_{RT} given $\pi_{RT}(\sigma) = 1/n!$ is reversible for Q_{RT}.

Theorem 1.2 (Ricci bound for the random transposition model)**.** *Let $n > 1$. The Ricci curvature of the random transposition model $(\mathcal{S}_n, Q_{\mathrm{RT}}, \pi_{\mathrm{RT}})$ is bounded from below by $\frac{4}{n(n-1)}$.*

As mentioned above the mixing time for $\mathcal{S}_n$ has been obtained by Diaconis and Shahshahani in [7]. The *coarse Ricci curvature* of the random transposition model can be estimated from above and below in a straightforward manner using contraction of the W_1-transportation distance (as observed by Gozlan et al [14], while very likely in the folklore) and shown to be of order n^{-2}. The modified logarithmic Sobolev inequality was studied by Goel [13], Gao–Quastel [11] and Bobkov–Tetali [3], who proved that

$$\frac{1}{n-1} \;\leq\; \alpha \;\leq\; \frac{4}{n-1}\;,$$

where the upper bound comes from the known spectral gap $\lambda = \frac{2}{n-1}$. Thus α and λ are both of order n^{-1}. Combining this estimate with Theorem 1.2, we infer that $4/(n^2 - n) \leq \kappa \leq 2/(n-1)$. It remains an open question to determine the correct order.

2. Preliminaries on Ricci curvature

We briefly recall some preliminaries on the notion of Ricci curvature for finite Markov chains following [18, 9, 10].

2.1. **Ricci curvature for Markov triples.** Let L be the generator of a continuous time Markov chain on a finite set $\mathcal{X}$. Thus the action of L on functions $\psi : \mathcal{X} \to \mathbb{R}$ is given by

$$L\psi(x) := \sum_{y \in \mathcal{X}} Q(x,y)(\psi(y) - \psi(x)) \ , \qquad x \in \mathcal{X} \ ,$$

with $Q(x,y) \geq 0$ for all $x \neq y$. Let π be a reversible measure for L, i.e. the detailed balance conditions

$$Q(x,y)\pi(x) = Q(y,x)\pi(y)$$

hold for all $x \neq y$. We refer to the triple $(\mathcal{X}, Q, \pi)$ as a Markov triple.

Let

$$\mathscr{P}(\mathcal{X}) := \Big\{ \rho : \mathcal{X} \to \mathbb{R}_+ \mid \sum_{x \in \mathcal{X}} \pi(x)\rho(x) = 1 \Big\}$$

be the set of *probability densities* (with respect to π) on $\mathcal{X}$. The subset consisting of those probability densities that are strictly positive is denoted by $\mathscr{P}_*(\mathcal{X})$.

A crucial role in this paper is played by the nonlocal transport metric $\mathcal{W}$ on $\mathscr{P}(\mathcal{X})$, which was introduced in [18, 19] (see also [5] for closely related metrics). In several ways, this metric can be regarded as a natural discrete analogue of the 2-Wasserstein metric [12]. The definition is based on a discrete analogue of the Benamou-Brenier formula: for $\rho_0, \rho_1 \in \mathscr{P}(\mathcal{X})$ we set

$$\mathcal{W}(\rho_0,\rho_1)^2 := \inf_{\rho,\psi} \left\{ \frac{1}{2} \int_0^1 \sum_{x,y \in \mathcal{X}} (\psi_t(x) - \psi_t(y))^2 \hat{\rho}_t(x,y) Q(x,y)\pi(x) \,\mathrm{d}t \right\} ,$$

where the infimum runs over all piecewise smooth curves $\rho : [0,1] \to \mathscr{P}(\mathcal{X})$ and all $\psi : [0,1] \times \mathcal{X} \to \mathbb{R}$ satisfying the discrete "continuity equation"

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}\rho_t(x) + \displaystyle\sum_{y \in \mathcal{X}} (\psi_t(y) - \psi_t(x))\hat{\rho}_t(x,y)Q(x,y) \ = \ 0 \ , \qquad x \in \mathcal{X} \ , \\ \rho(0) = \rho_0 \ , \qquad \rho(1) = \rho_1 \ . \end{cases} \tag{2.1}$$

Here, given $\rho \in \mathscr{P}(\mathcal{X})$, we write $\hat{\rho}(x,y) := \theta\big(\rho(x), \rho(y)\big)$, where $\theta(r,s) = \int_0^1 r^{1-p}s^p \,\mathrm{d}p$ is the logarithmic mean of r and s.

The relative entropy (with respect to π) of $\rho \in \mathscr{P}(\mathcal{X})$ is defined as usual by

$$\mathcal{H}(\rho) = \sum_{x \in \mathcal{X}} \pi(x)\rho(x) \log \rho(x) \ . \tag{2.2}$$

It turns out that the metric $\mathcal{W}$ is induced by a Riemannian structure on the interior $\mathscr{P}_*(\mathcal{X})$ of $\mathscr{P}(\mathcal{X})$. Moreover, every pair of densities $\rho_0, \rho_1 \in \mathscr{P}(\mathcal{X})$ can be joined by a constant speed geodesic, i.e., there exists a curve $\rho : [0,1] \to \mathscr{P}(\mathcal{X})$ connecting ρ_0 and ρ_1 satisfying $\mathcal{W}(\rho_s, \rho_t) = |t-s|\mathcal{W}(\rho_0,\rho_1)$ for all $s,t \in [0,1]$. Therefore, the following definition in the spirit of Lott–Sturm–Villani [17, 24] is meaningful.

Definition 2.1 (Discrete Ricci curvature)**.** *We say that a Markov triple* $(\mathcal{X}, Q, \pi)$ *has* Ricci curvature bounded from below by $\kappa \in \mathbb{R}$ *if for any constant speed geodesic* $\{\rho_t\}_{t \in [0,1]}$ *in* $(\mathscr{P}(\mathcal{X}), \mathcal{W})$ *we have*

$$\mathcal{H}(\rho_t) \leq (1-t)\mathcal{H}(\rho_0) + t\mathcal{H}(\rho_1) - \frac{\kappa}{2} t(1-t)\mathcal{W}(\rho_0,\rho_1)^2 \ .$$

In this case, we write $\mathrm{Ric}(\mathcal{X}, Q, \pi) \geq \kappa$*, or simply* $\mathrm{Ric}(Q) \geq \kappa$.

2.2. **Equivalent conditions for Ricci curvature.** To proceed further, we introduce the following convenient notation. For a function $\varphi : \mathcal{X} \to \mathbb{R}$ we consider the *discrete gradient* $\nabla\varphi \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ defined by

$$\nabla\varphi(x,y) := \varphi(y) - \varphi(x) \ .$$

For $\Psi \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ we consider the *discrete divergence* $\nabla \cdot \Psi \in \mathbb{R}^{\mathcal{X}}$ defined by

$$(\nabla\cdot\Psi)(x) := \frac{1}{2}\sum_{y\in\mathcal{X}}(\Psi(x,y) - \Psi(y,x))Q(x,y) \in \mathbb{R} \ .$$

With this notation we have $L := \nabla\cdot\nabla$, and the integration by parts formula

$$\langle \nabla\psi, \Psi\rangle_\pi = -\langle \psi, \nabla\cdot\Psi\rangle_\pi$$

holds. Here we write, for $\varphi, \psi \in \mathbb{R}^{\mathcal{X}}$ and $\Phi, \Psi \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$,

$$\begin{aligned}
\langle\varphi,\psi\rangle_\pi &= \sum_{x\in\mathcal{X}} \varphi(x)\psi(x)\pi(x) \ , \\
\langle\Phi,\Psi\rangle_\pi &= \frac{1}{2}\sum_{x,y\in\mathcal{X}} \Phi(x,y)\Psi(x,y)Q(x,y)\pi(x) \ .
\end{aligned}$$

An important role in our analysis is played by the quantity $\mathcal{B}(\rho,\psi)$, which is defined for $\rho \in \mathbb{R}_+^{\mathcal{X}}$ and $\psi \in \mathbb{R}^{\mathcal{X}}$ by

$$\begin{aligned}
\mathcal{B}(\rho,\psi) :=& \frac{1}{2}\big\langle \widehat{L}\rho \cdot \nabla\psi, \nabla\psi\big\rangle_\pi - \big\langle \hat\rho\cdot\nabla\psi \ , \nabla L\psi\big\rangle_\pi \\
=& \frac{1}{4}\sum_{x,y,z\in\mathcal{X}} \big(\psi(x)-\psi(y)\big)^2\Big(\hat\rho_1(x,y)\big(\rho(z)-\rho(x)\big)Q(x,z) \\
&\qquad + \hat\rho_2(x,y)\big(\rho(z)-\rho(y)\big)Q(y,z)\Big)Q(x,y)\pi(x) \\
&- \frac{1}{2}\sum_{x,y,z\in\mathcal{X}} \Big(Q(x,z)\big(\psi(z)-\psi(x)\big) - Q(y,z)\big(\psi(z)-\psi(y)\big)\Big) \\
&\qquad \times \big(\psi(x)-\psi(y)\big)\hat\rho(x,y)Q(x,y)\pi(x) \ ,
\end{aligned} \tag{2.3}$$

where

$$\begin{aligned}
\hat\rho(x,y) &:= \theta(\rho(x),\rho(y)) \ , \\
\hat\rho_i(x,y) &:= \partial_i\theta(\rho(x),\rho(y)) \ , \quad i = 1,2 \ , \\
\widehat{L}\rho(x,y) &:= \hat\rho_1(x,y)L\rho(x) + \hat\rho_2(x,y)L\rho(y) \ .
\end{aligned}$$

The term $\mathcal{B}(\rho,\psi)$ is reminiscent of the Bochner formula in Riemannian geometry, which asserts that $\frac{1}{2}\Delta|\nabla\psi|^2 - \langle\nabla\Delta\psi,\nabla\psi\rangle = \mathrm{Ric}(\nabla\psi,\nabla\psi) + \|D^2\psi\|_{HS}^2$.

Let us further introduce the quantity

$$\mathcal{A}(\rho,\psi) := \langle\hat\rho\cdot\nabla\psi,\nabla\psi\rangle = \frac{1}{2}\sum_{x,y\in\mathcal{X}}(\psi(y)-\psi(x))^2\hat\rho(x,y)Q(x,y)\pi(x) \ ,$$

for $\rho \in \mathscr{P}(\mathcal{X})$ and $\psi \in \mathbb{R}^{\mathcal{X}}$.

The following result from [9] provides a reformulation of Ricci lower bounds in terms of $\mathcal{B}$ and $\mathcal{A}$.

Theorem 2.2 (Characterisation of Ricci curvature bounds)**.** *Let $\kappa \in \mathbb{R}$. For an irreducible and reversible Markov kernel $(\mathcal{X}, Q, \pi)$ the following assertions are equivalent:*

(1) $\mathrm{Ric}(\mathcal{X}, Q, \pi) \geq \kappa$ *;*

(2) For all $\rho \in \mathscr{P}_*(\mathcal{X})$ we have

$$\mathrm{Hess}\,\mathcal{H}(\rho) \geq \kappa \ ;$$

(3) For all $\rho \in \mathscr{P}_*(\mathcal{X})$ and $\psi \in \mathbb{R}^{\mathcal{X}}$ we have

$$\mathcal{B}(\rho, \psi) \geq \kappa \mathcal{A}(\rho, \psi) \ .$$

The equivalence between (1) and (2) shows equivalence of a non-smooth and a smooth notion of convexity. This equivalence is non-trivial, since the Riemannian metric is degenerate at the boundary. Assertion (3) is an explicit reformulation of (2). The inequality in (3) can be seen as a discrete analogue of Bochner's inequality.

3. A simple criterion for Ricci curvature bounds

Here we present a combinatorial method for controlling the quantity $\mathcal{B}$ from (2.3). We will first study this quantity in detail in the case where the Markov chain is simple random walk on a triangle or on a square. The resulting bounds will then be applied to concrete examples with sufficient symmetry in which the underlying graph can be decomposed into squares and triangles.

3.1. **Decomposition of $\mathcal{B}(\rho, \psi)$.**

Let us consider the natural graph structure $(\mathcal{X}, E)$ associated with the kernel Q, where the set of edges is defined by

$$E \ := \big\{ \{x, y\} \ : \ Q(x, y) > 0 \big\} \ .$$

Then we can rewrite the quantity $\mathcal{A}$ as

$$\mathcal{A}(\rho, \psi) \ = \ \sum_{e \in E} a(e) c(e) \ ,$$

where for $e = \{x, y\}$ we set $c(e) = Q(x, y)\pi(x)$ and

$$a(e) \ = \ \big(\psi(y) - \psi(x)\big)^2 \hat{\rho}(x, y) \ .$$

Given two edges $e, e' \in E$ we write $e \sim e'$ iff they are adjacent or identical, i.e., iff $e = \{x, y\}, e' = \{x, z\}$ for some $x, y, z \in \mathcal{X}$. Then we can rewrite the quantity $\mathcal{B}$ as a sum over pairs of adjacent edges. It will be convenient to write $c(x, y) := Q(x, y)\pi(x)$. Note that the reversibility assumption implies that $c(x, y) = c(y, x)$.

Lemma 3.1 (Reformulation of $\mathcal{B}(\rho, \psi)$)**.** *For all $\rho \in \mathbb{R}_+^{\mathcal{X}}$ and $\psi \in \mathbb{R}^{\mathcal{X}}$ we have*

$$\mathcal{B}(\rho, \psi) \ = \ \sum_{e, e' \in E, e \sim e'} b(e, e') \ , \tag{3.1}$$

where for $e = \{x, y\}$ and $e' = \{x, z\}$ with $y \neq z$ we set

$$\begin{aligned} b(e, e') \ := \ \Big[&\frac{1}{2} \big(\psi(x) - \psi(y)\big)^2 \hat{\rho}_1(x, y) \big(\rho(z) - \rho(x)\big) \\ &+ \big(\psi(y) - \psi(x)\big)\big(\psi(z) - \psi(x)\big) \hat{\rho}(x, y) \Big] Q(x, z) c(x, y) \ , \end{aligned}$$

while for $y = z$ we set

$$b(e,e) \;:=\; \frac{1}{2}\big(\psi(x)-\psi(y)\big)^2\Big[2\hat\rho(x,y)\big[Q(x,y)+Q(y,x)\big] \\ + \hat\rho_1(x,y)\big(\rho(y)-\rho(x)\big)Q(x,y)+\hat\rho_2(x,y)\big(\rho(x)-\rho(y)\big)Q(y,x)\Big]c(x,y)\;.$$

Proof. First note that using the fact that $\hat\rho(x,y)=\hat\rho(y,x)$, $\hat\rho_1(x,y)=\hat\rho_2(y,x)$ and the detailed balance condition $Q(x,y)\pi(x)=Q(y,x)\pi(y)$ we can rewrite (2.3) in the form

$$\mathcal{B}(\rho,\psi) \;=\; \frac{1}{2}\sum_{x,y,z\in\mathcal{X}}\big(\psi(x)-\psi(y)\big)^2\hat\rho_1(x,y)\big(\rho(z)-\rho(x)\big)Q(x,z)Q(x,y)\pi(x) \\ + \sum_{x,y,z\in\mathcal{X}}\big(\psi(y)-\psi(x)\big)\big(\psi(z)-\psi(x)\big)\hat\rho(x,y)Q(x,z)Q(x,y)\pi(x)\;.$$

Now the assertion is obvious. □

Given a subgraph $G=(\mathcal{Y},F)$ of $(\mathcal{X},E)$ and two functions $\rho\in\mathbb{R}_+^{\mathcal{X}}$ and $\psi\in\mathbb{R}^{\mathcal{X}}$ we denote by ρ^G,ψ^G their restrictions to $\mathcal{Y}$. Moreover, we set $\mathcal{A}_G(\rho,\psi)$ and $\mathcal{B}_G(\rho,\psi)$ to be the quantities $\mathcal{A},\mathcal{B}$ calculated in the weighted graph $(\mathcal{Y},F)$ with the functions ρ^G,ψ^G. More precisely,

$$\mathcal{A}_G(\rho,\psi) \;:=\; \sum_{e\in F} a(e)c(e)\;,$$

$$\mathcal{B}_G(\rho,\psi) \;:=\; \sum_{e,e'\in F, e\sim e'} b(e,e')\;.$$

Further, it turns out to be useful to seperate the contribution to $\mathcal{B}$ coming from identical edges ("on-diagonal entries") and from adjacent edges ("off-diagonal entries"). Thus we set

$$\mathcal{B}_G^{\mathrm{on}}(\rho,\psi) \;:=\; \sum_{e\in F} b(e,e)\;,$$

$$\mathcal{B}_G^{\mathrm{off}}(\rho,\psi) \;:=\; \sum_{e,e'\in F, e\sim e', e\neq e'} b(e,e')\;.$$

3.2. **An on-diagonal bound for d-regular graphs.** From now on let us assume that the Markov chain is simple random walk on a d-regular graph $(\mathcal{X},E)$, i.e.,

$$Q(x,y) \;=\; \begin{cases} \frac{1}{d}\;, & \{x,y\}\in E\;,\\ 0\;, & \text{otherwise}\;.\end{cases}$$

The uniform probability measure given by $\pi(x)=\mu=|\mathcal{X}|^{-1}$ for all $x\in\mathcal{X}$ satisfies the detailed balance condition.

The following results is a general bound on the on-diagonal part of $\mathcal{B}$. In the proof we shall use the following elementary properties of the logarithmic mean:

$$s\partial_1\theta(s,t)+t\partial_2(s,t) \;=\; \theta(s,t)\;, \tag{3.2}$$

$$u\partial_1\theta(s,t)+v\partial_2(s,t) \;\geq\; \theta(u,v)\;, \tag{3.3}$$

for all $s,t,u,v>0$. A proof can be found in [9, Lemma 2.2].

Lemma 3.2 (On-diagonal bound)**.** *For every subgraph $G \subset (\mathcal{X}, E)$ and all $\rho \in \mathbb{R}_+^{\mathcal{X}}$ and $\psi \in \mathbb{R}^{\mathcal{X}}$ we have*

$$\mathcal{B}_G^{\rm on}(\rho,\psi) \;\geq\; \frac{2}{d}\mathcal{A}_G(\rho,\psi)\;. \tag{3.4}$$

Proof. Let us write $G = (Y, F)$. Using (3.2) and (3.3) we obtain

$$\begin{aligned}
&\mathcal{B}_G^{\rm on}(\rho,\psi)\\
&= \frac{\mu}{d^2}\sum_{\{x,y\}\in F}\frac{1}{2}\Big(\psi(y)-\psi(x)\Big)^2\Big[\hat\rho_1(x,y)\big(\rho(y)-\rho(x)\big)+\hat\rho_2(x,y)\big(\rho(x)-\rho(y)\big)+4\hat\rho(x,y)\Big]\\
&\geq \frac{2\mu}{d^2}\sum_{\{x,y\}\in F}\Big(\psi(y)-\psi(x)\Big)^2\hat\rho(x,y) \;=\; \frac{2}{d}\mathcal{A}_G(\rho,\psi)\;,
\end{aligned}$$

which is the desired bound. □

3.3. **Off-diagonal bounds for triangles and squares.** For the off-diagonal part, let us first consider the special cases where the subgraph G is a triangle or a square.

Lemma 3.3 (Off-diagonal bound for triangles)**.** *Let $\triangle = (\mathcal{Y}, F)$ be a triangle subgraph of $(\mathcal{X}, E)$, i.e., $\mathcal{Y} = \{x_1, x_2, x_3\}$ and $F = \big\{\{x_i, x_{i+1}\},\ i = 1, 2, 3\big\}$ for some distinct $x_i \in \mathcal{X}$. Then, for any $\rho \in \mathbb{R}_+^{\mathcal{X}}$ and $\psi \in \mathbb{R}^{\mathcal{X}}$ we have*

$$\mathcal{B}_\triangle^{\rm off}(\rho,\psi) \;\geq\; \frac{1}{2d}\mathcal{A}_\triangle(\rho,\psi)\;. \tag{3.5}$$

Proof. For convenience we set $\rho_i = \rho(x_i)$ and $g_i = \psi(x_{i+1}) - \psi(x_i)$ for $i = 1, 2, 3$ with the convention that $x_0 = x_3$ and $x_4 = x_1$. To simplify notation we write $\hat\rho_{i,j} = \hat\rho(x_i, x_j)$, $\hat\rho^1_{i,j} = \hat\rho_1(x_i, x_j)$ and $\hat\rho^2_{i,j} = \hat\rho_2(x_i, x_j)$. It is readily verified that

$$\mathcal{B}_\triangle^{\rm off}(\rho,\psi) \;=\; \frac{\mu}{d^2}\sum_{i=1}^{3}\frac{1}{2}g_i^2\Big[\hat\rho^1_{i,i+1}(\rho_{i-1}-\rho_i)+\hat\rho^2_{i,i+1}(\rho_{i+2}-\rho_{i+1})\Big]-g_i(g_{i+1}+g_{i-1})\hat\rho_{i,i+1}\;.$$

Using the inequality $\hat\rho^1_{ij} \geq 0$, the identity (3.2), and the fact that $g_1 + g_2 + g_3 = 0$, we estimate

$$\begin{aligned}
\mathcal{B}_\triangle^{\rm off}(\rho,\psi) \;&\geq\; \frac{\mu}{d^2}\sum_{i=1}^{3}-\frac{1}{2}g_i^2\hat\rho_{i,i+1}-g_i(g_{i+1}+g_{i-1})\hat\rho_{i,i+1}\\
&= \frac{\mu}{2d^2}\sum_{i=1}^{3}g_i^2\hat\rho_{i,i+1} \;=\; \frac{1}{2d}\mathcal{A}_\triangle(\rho,\psi)\;,
\end{aligned}$$

which completes the proof. □

For $s, t, u, v > 0$ let $\mathsf{D}(s, t; u, v) := u\partial_1\theta(s, t) + v\partial_2(s, t) - \theta(u, v)$ be the deficit in the 4-point inequality (3.3), thus $\mathsf{D}(s, t; u, v) \geq 0$. To simplify notation we will often write $\mathsf{D}^{k,l}_{i,j}$ instead of $\mathsf{D}(\rho_i, \rho_j; \rho_k, \rho_l)$. The following result provides a convenient representation of $\mathcal{B}^{\rm off}_\square$ as a sum of nonnegative terms.

Lemma 3.4 (Off-diagonal bound for squares)**.** *Let $\square = (\mathcal{Y}, F)$ be a square subgraph of $(\mathcal{X}, E)$, i.e., $\mathcal{Y} = \{x_1, x_2, x_3, x_4\}$ and $F = \big\{\{x_i, x_{i+1}\},\ i = 1, \cdots, 4\big\}$ for some distinct $x_i \in \mathcal{X}$. Then,* *for any $\rho \in \mathbb{R}_+^{\mathcal{X}}$ and $\psi \in \mathbb{R}^{\mathcal{X}}$ we have*

$$\mathcal{B}_\square^{\rm off}(\rho,\psi) = \frac{\mu}{2d^2}|AS|^2(\psi;\square)\sum_{i=1}^4 \hat\rho(x_i,x_{i+1}) + \frac{\mu}{2d^2}\sum_{i=1}^4 \big(\psi(x_{i+1})-\psi(x_i)\big)^2 \mathsf{D}_{i,i+1}^{i-1,i+2} \;\geq\; 0\;,$$

where $|AS|(\psi;\square) := |\psi(x_1)-\psi(x_2)+\psi(x_3)-\psi(x_4)|$ denotes the alternating sum of ψ on $\square$.

Note that the definition of $|AS|(\psi;\square)$ does not depend on the parametrisation of $\square$.

Proof. We set $\rho_i = \rho(x_i)$ and $g_i = \psi(x_{i+1})-\psi(x_i)$ for $i=1,2,3,4$ with the convention $x_0 = x_4$ and $x_5 = x_1$. Moreover, we define $\hat\rho_{i,j}$ and $\hat\rho^1_{i,j}, \hat\rho^2_{i,j}$ as in the proof of Lemma 3.3. Using (3.2), (3.3) and the identity $g_1+g_2+g_3+g_4 = 0$ we obtain

$$\begin{aligned}
\mathcal{B}_\square^{\rm off}(\rho,\psi) &= \frac{\mu}{d^2}\sum_{i=1}^4 \frac12 g_i^2\Big[\hat\rho^1_{i,i+1}(\rho_{i-1}-\rho_i) + \hat\rho^2_{i,i+1}(\rho_{i+2}-\rho_{i+1})\Big] - g_i(g_{i+1}+g_{i-1})\hat\rho_{i,i+1}\\
&= \frac{\mu}{d^2}\sum_{i=1}^4 \frac12 g_i^2\Big[\hat\rho^1_{i,i+1}\rho_{i-1} + \hat\rho^2_{i,i+1}\rho_{i+2}\Big] - g_i(g_{i+1}+\frac12 g_i+g_{i-1})\hat\rho_{i,i+1}\\
&= \frac{\mu}{d^2}\sum_{i=1}^4 \frac12 g_i^2\Big[\hat\rho_{i-1,i+2} + \mathsf{D}_{i,i+1}^{i-1,i+2}\Big] + g_i(\frac12 g_i+g_{i+2})\hat\rho_{i,i+1}\\
&= \frac{\mu}{4d^2}\sum_{i=1}^4 (g_i+g_{i+2})^2\Big[\hat\rho_{i-1,i+2}+\hat\rho_{i,i+1}\Big] + 2g_i^2\mathsf{D}_{i,i+1}^{i-1,i+2}\\
&= \frac{\mu}{8d^2}\sum_{i=1}^4 (g_i+g_{i+2})^2\Big[\sum_{j=1}^4 \hat\rho_{j,j+1}\Big] + 4g_i^2\mathsf{D}_{i,i+1}^{i-1,i+2}\;,
\end{aligned}$$

which yields the desired identity. □

Remark 3.5. The bound $\mathcal{B}_\square^{\rm off}(\rho,\psi)\geq 0$ is sharp, in the sense that there exist $\rho\in\mathbb{R}_+^{\mathcal{X}}$ and $\psi\in\mathbb{R}^{\mathcal{X}}$ with $\mathcal{B}_\square^{\rm off}(\rho,\psi) = 0$ and $\mathcal{A}(\rho,\psi) > 0$. Take for instance $\rho_i = 1$ for all i and a non-nonstant function ψ with $|AS|(\psi;\square) = 0$.

4. The Bernoulli–Laplace model

For integers $n>1$ and $1\leq k\leq n-1$ consider the k-slice of the n-dimensional discrete cube

$$\Omega(n,k) = \big\{x\in\{0,1\}^n \;:\; x_1+\cdots+x_n = k\big\}\;.$$

Two points in $\Omega(n,k)$ are declared neighbors if they differ in exactly two coordinates. Let us set

$$I(x) = \{i\leq n \;:\; x_i = 1\}\;,\quad J(x) = \{j\leq n \;:\; x_j = 0\}\;.$$

Then the neighbors of x are given by $\{s_{ij}x\}_{i\in I(x), j\in J(x)}$, where

$$(s_{ij}x)_i = 0\;,\ (s_{ij}x)_j = 1\;, (s_{ij}x)_k = x_k \quad \forall k\neq i,j$$

Note that every point $x\in\Omega(n,k)$ has $k(n-k)$ neighbors, and that the set of edges is $E = \big\{\{x,s_{ij}x\} \;:\; x\in\Omega(n,k),\ i\in I(x),\ j\in J(x)\big\}$. The simple random walk on $(\Omega(n,k),E)$ is given by $Q_{\rm BL}(x,y) = (k(n-k))^{-1}$ whenever $x\sim y$, and has as invariant measure the uniform measure $\pi(x) = |\Omega(n,k)|^{-1} = \binom{n}{k}^{-1}$.

We have the following curvature bound for the Bernoulli Laplace model (Theorem 1.1).

Theorem 4.1. *The simple random walk Q_{BL} on $\Omega(n,k)$ satisfies*

$$\mathrm{Ric}(Q_{\mathrm{BL}}) \geq \frac{n+2}{2k(n-k)} .$$

Proof. Let us set $d = k(n-k)$ and $\mu = \binom{n}{k}^{-1}$. Then we need to show that for any $\rho \in \mathscr{P}_*(\Omega(n,k))$ and any $\psi : \Omega(n,k) \to \mathbb{R}$ we have

$$\mathcal{B}(\rho,\psi) \ \geq \ \frac{n+2}{2d}\mathcal{A}(\rho,\psi) .$$

Let $P = \big\{(e,e') \in E \times E \ : \ e \sim e', e \neq e'\big\}$ be the set of pairs of adjacent non-identical edges. We define a decomposition $P = P_1 \cup P_2$ as follows. For $(e,e') \in P$ we have $e = \{x, s_{ij}x\}$ and $e' = \{x, s_{pq}x\}$ for some $x \in \Omega(n,k)$ and $i,p \in I(x)$, $j,q \in J(x)$. We say that $(e,e') \in P_1$ if e,e' "overlap", i.e., $i = p$ or $j = q$. Otherwise, if $i \neq p$ and $j \neq q$ we say that $(e,e') \in P_2$. Now we can write

$$\begin{aligned} \mathcal{B}(\rho,\psi) \ &= \ \mathcal{B}^{\mathrm{on}}(\rho,\psi) + \mathcal{B}^{\mathrm{off},1}(\rho,\psi) + \mathcal{B}^{\mathrm{off},2}(\rho,\psi) , \qquad \text{where} \\ \mathcal{B}^{\mathrm{off},i}(\rho,\psi) \ &= \ \sum_{(e,e') \in P_i} b(e,e') , \quad i = 1,2 . \end{aligned}$$

Note that every pair $(e,e') \in P_1$ is part of a unique triangle in the graph $(\Omega(n,k),E)$. Indeed, $s_{ij}x$ and $s_{iq}x$ differ in exactly two coordinates, namely j and q. Similarly, $s_{ij}x$ and $s_{pj}x$ differ exactly in i and p. Moreover, every edge $e \in E$ is part of $n-2$ different triangles. Indeed, any two neighbors $x, s_{ij}x$ have exactly $n-2$ common neighbors, namely the points $s_{iq}x$, $q \in J(x) \setminus \{j\}$ and $s_{pj}x$, $p \in I(x) \setminus \{i\}$. Thus we obtain

$$\mathcal{B}^{\mathrm{off},1}(\rho,\psi) \ = \ \sum_{\triangle} \mathcal{B}^{\mathrm{off}}_{\triangle}(\rho,\psi) \ \geq \ \frac{1}{2d}\sum_{\triangle} \mathcal{A}_{\triangle}(\rho,\psi) \ = \ \frac{n-2}{2d}\mathcal{A}(\rho,\psi) ,$$

where we haved summed over all triangle subgraphs $\triangle$ and used Lemma 3.3. Now note that every pair $(e,e') \in P_2$ is part of precisely two squares in the graph $(\Omega(n,k),E)$. Indeed, if $i \neq p$ and $j \neq q$, the points $x, s_{ij}x, s_{pq}s_{ij}x, s_{pq}x$ and the points $x, s_{ij}x, s_{iq}x, s_{pq}x$ form a cycle. Thus, using Lemma 3.4, we obtain

$$\mathcal{B}^{\mathrm{off},2}(\rho,\psi) \ = \ \frac{1}{2}\sum_{\square} \mathcal{B}^{\mathrm{off}}_{\square}(\rho,\psi) \ \geq \ 0 ,$$

where we have summed over all square subgraphs $\square$.

Finally, putting everything together and using Lemma 3.2 we get

$$\mathcal{B}(\rho,\psi) \ \geq \ \frac{2}{d}\mathcal{A}(\rho,\psi) + \frac{n-2}{2d}\mathcal{A}(\rho,\psi) \ = \ \frac{n+2}{2d}\mathcal{A}(\rho,\psi) .$$

□

As noted in the introduction, we recover the best known constant in the modified logarithmic Sobolev inequality as a corollary.

5. The random transposition model

Let $\mathcal{S}_n$ be the set of permutations of $[n] := \{1, \ldots, n\}$, i.e., $\mathcal{S}_n$ consists of all bijective maps $\sigma : [n] \to [n]$. The composition $\sigma_1 \circ \sigma_2$ of two permutations $\sigma_1, \sigma_2 \in \mathcal{S}_n$ will be denoted by $\sigma_1\sigma_2$. For $1 \le i < j \le n$ let $\tau_{ij} \in \mathcal{S}_n$ denote the transposition which interchanges i and j, i.e.,

$$\tau_{ij}(i) = j\ , \quad \tau_{ij}(j) = i\ , \quad \tau_{ij}(k) = k \quad \forall k \neq i,j\ .$$

We define a graph structure on the group $\mathcal{S}_n$ by saying that two permutations are neighbors if they differ by precisely one transposition. Thus every vertex $\sigma \in \mathcal{S}_n$ has $\binom{n}{2}$ neighbors given by $\{\tau_{ij}\sigma\}_{1\le i<j\le n}$, and the set of edges is $E = \big\{\{\sigma, \tau_{ij}\sigma\} \ : \ 1 \le i < j \le n\big\}$. The simple random walk on $(\mathcal{S}_n, E_n)$ is given by the Markov transition rates

$$Q_{\mathrm{RT}}(\sigma, \eta) = \begin{cases} \frac{2}{n(n-1)}\ , & \text{if } \sigma \sim \eta\ , \\ 0\ , & \text{otherwise}\ , \end{cases}$$

and the uniform probability measure π_n given by $\pi_n(\sigma) = |\mathcal{S}_n|^{-1} = (n!)^{-1}$ is reversible for Q.

We have the following curvature bound for the random transposition model.

Theorem 5.1. *The simple random walk Q_{RT} on $\mathcal{S}_n$ satisfies*

$$\mathrm{Ric}(Q_{\mathrm{RT}}) \ge \frac{4}{n(n-1)}\ .$$

Proof. Let us set $d = \frac{n(n-1)}{2}$. Then we need to show that for any $\rho \in \mathscr{P}_*(\mathcal{S}_n)$ and $\psi \in \mathbb{R}^{\mathcal{X}}$ we have

$$\mathcal{B}(\rho, \psi) \ \ge \ \frac{2}{d}\mathcal{A}(\rho, \psi)\ .$$

Let $P = \big\{(e, e') \in E \times E \ : \ e \sim e', e \neq e'\big\}$ be the set of pairs of adjacent non-identical edges. We define a decomposition $P = P_1 \cup P_2$ as follows. For $(e, e') \in P$ we have $e = \{\sigma, \tau_{ij}\sigma\}$ and $e' = \{\sigma, \tau_{pq}x\}$ for some $\sigma \in \mathcal{S}_n$ and $i < j, p < q$. We say that $(e, e') \in P_1$ if e, e' do not "overlap", i.e., $\{i, j\} \cap \{p, q\} = \emptyset$. Otherwise, if $\{i, j\} \cap \{p, q\} \neq \emptyset$ we say $(e, e') \in P_2$. Now we can write

$$\begin{aligned} \mathcal{B}(\rho, \psi) \ &= \ \mathcal{B}^{\mathrm{on}}(\rho, \psi) + \mathcal{B}^{\mathrm{off},1}(\rho, \psi) + \mathcal{B}^{\mathrm{off},2}(\rho, \psi)\ , \qquad \text{where} \\ \mathcal{B}^{\mathrm{off},i}(\rho, \psi) \ &= \ \sum_{(e,e')\in P_i} b(e, e')\ , \quad i = 1, 2\ . \end{aligned}$$

Note that every pair $(e, e') \in P_1$ is part of a unique square in the graph $(\mathcal{S}_n, E)$. Indeed, $\tau_{ij}\sigma$ and $\tau_{pq}\sigma$ have the vertex $\tau_{pq}\tau_{ij}\sigma = \tau_{ij}\tau_{pq}\sigma$ as their unique common neighbor besides σ. Observe that all pairs of adjacent edges in this square belong to P_1. Every pair $(e, e') \in P_2$ is part of exactly two squares. Indeed, let $(e, e') \in P_2$, and assume without loss of generality that $e = \{\sigma, \tau_{ij}\sigma\}$ and $e' = \{\sigma, \tau_{iq}\sigma\}$. Then $\tau_{iq}\tau_{ij}\sigma = \tau_{jq}\tau_{iq}\sigma$ and $\tau_{jq}\tau_{ij}\sigma = \tau_{ij}\tau_{iq}\sigma$ are the two distinct common neighbors of $\tau_{ij}\sigma$ and $\tau_{iq}\sigma$. Note that all pairs of adjacent edges in these squares belong to P_2.

For $i = 1, 2$, we let A_i denote the set of all square subgraphs of $(\mathcal{S}_n, E)$ in which each adjacent pair of edges belongs to P_i. Thus we obtain

$$\mathcal{B}^{\mathrm{off},1}(\rho,\psi) \;=\; \sum_{\square\in A_1} \mathcal{B}^{\mathrm{off}}_{\square}(\rho,\psi) \;\geq\; 0\;,$$

$$\mathcal{B}^{\mathrm{off},2}(\rho,\psi) \;=\; \frac{1}{2}\sum_{\square\in A_2} \mathcal{B}^{\mathrm{off}}_{\square}(\rho,\psi) \;\geq\; 0\;.$$

Finally, putting everything together and using Lemma 3.2 we get

$$\mathcal{B}(\rho,\psi) \;\geq\; \mathcal{B}^{\mathrm{on}}(\rho,\psi) \;\geq\; \frac{2}{d}\mathcal{A}(\rho,\psi)\;,$$

which completes the proof. $\square$

One might hope that the above lower Ricci bound can be improved by looking at subgraphs isomorphic to $\mathcal{S}_3$ taking over the role of the triangles in the Bernoulli–Laplace model. However, the next lemma shows that they only give a nonnegative contribution to the off-diagonal $\mathcal{B}$-term in general, which does not improve the bound obtained in Theorem 5.1.

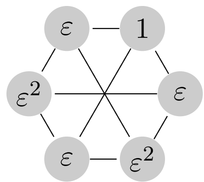
Lemma 5.2. *For all $\rho : \mathcal{S}_3 \to \mathbb{R}_+$ and $\psi : \mathcal{S}_3 \to \mathbb{R}$ we have*

$$\mathcal{B}^{\mathrm{off}}(\rho,\psi) \;\geq\; 0\;.$$

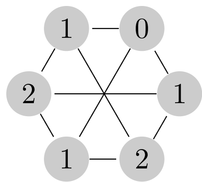
Moreover, this bound is sharp in the sense that for any $\kappa > 0$ there exist ρ and ψ such that $\mathcal{B}^{\mathrm{off}}(\rho,\psi) < \kappa\mathcal{A}(\rho,\psi)$.

Proof. The non-negativity follows from writing $\mathcal{B}^{\mathrm{off}}$ as a sum of contributions from squares and using Lemma 3.4. To see that this bound is sharp, define ρ_ε and ψ by

ρ_ε :



ψ :



Then one can check that as $\varepsilon \to 0$,

$$\frac{\mathcal{B}^{\mathrm{off}}(\rho_\varepsilon,\psi)}{\mathcal{A}(\rho_\varepsilon,\psi)} \;\to\; 0\;.$$

$\square$

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. *Duke Math. J.*, 163(7):1405–1490, 2014.

[2] D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985.

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

[4] 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.

[5] S.-N. Chow, W. Huang, Y. Li, and H. Zhou. Fokker-Planck equations for a free energy functional or Markov process on a graph. *Arch. Ration. Mech. Anal.*, 203(3):969–1008, 2012.

[6] P. Diaconis and L. Saloff-Coste. Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.*, 6(3):695–750, 1996.
[7] P. Diaconis and M. Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981.
[8] P. Diaconis and M. Shahshahani. Time to reach stationarity in the Bernoulli-Laplace diffusion model. *SIAM J. Math. Anal.*, 18(1):208–218, 1987.
[9] M. Erbar and J. Maas. Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.*, 206(3):997–1038, 2012.
[10] M. Erbar and J. Maas. Gradient flow structures for discrete porous medium equations. *Discrete Contin. Dyn. Syst.*, 34(4):1355–1374, 2014.
[11] 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.
[12] N. Gigli and J. Maas. Gromov-Hausdorff convergence of discrete transportation metrics. *SIAM J. Math. Anal.*, 45(2):879–899, 2013.
[13] S. Goel. Modified logarithmic Sobolev inequalities for some models of random walk. *Stochastic Process. Appl.*, 114(1):51–79, 2004.
[14] N. Gozlan, J. Melbourne, W. Perkins, C. Roberto, P-M. Samson, and P. Tetali. Working Group in New directions in mass transport: discrete versus continuous. *AIM SQuaRE report, October*, 2013.
[15] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Displacement convexity of entropy and related inequalities on graphs. *Probability Theory and Related Fields*, 160:47–94, 2014.
[16] T.-Y. Lee and H.-T. Yau. Logarithmic Sobolev inequality for some models of random walks. *Ann. Probab.*, 26(4):1855–1873, 1998.
[17] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. Math. (2)*, 169(3):903–991, 2009.
[18] J. Maas. Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.*, 261(8):2250–2292, 2011.
[19] A. Mielke. A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. *Nonlinearity*, 24(4):1329–1346, 2011.
[20] A. Mielke. Geodesic convexity of the relative entropy in reversible Markov chains. *Calc. Var. Partial Differential Equations, Online first*, 2012.
[21] Y. Ollivier. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.*, 256(3):810–864, 2009.
[22] Y. Ollivier. A survey of Ricci curvature for metric spaces and Markov chains. *Adv. Stud. Pure Math.*, 57:343–381, 2010.
[23] M.D. Sammer. Aspects of mass transportation in discrete concentration inequalities. *PhD thesis, Georgia Institute of Technology*, 2005.
[24] K.-Th. Sturm. On the geometry of metric measure spaces. I and II. *Acta Math.*, 196(1):65–177, 2006.

Matthias Erbar, Institute for Applied Mathematics, University of Bonn, Endenicher Allee 60, 53115 Bonn, Germany
E-mail address: erbar@iam.uni-bonn.de

Jan Maas, Institute for Applied Mathematics, University of Bonn, Endenicher Allee 60, 53115 Bonn, Germany
E-mail address: maas@uni-bonn.de

Prasad Tetali, School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332, USA
E-mail address: tetali@math.gatech.edu