RICCI CURVATURE OF FINITE MARKOV CHAINS VIA CONVEXITY OF THE ENTROPY

MATTHIAS ERBAR AND JAN MAAS

Abstract. We study a new notion of Ricci curvature that applies to Markov chains on discrete spaces. This notion relies on geodesic convexity of the entropy and is analogous to the one introduced by Lott, Sturm, and Villani for geodesic measure spaces. In order to apply to the discrete setting, the role of the Wasserstein metric is taken over by a different metric, having the property that continuous time Markov chains are gradient flows of the entropy.

Using this notion of Ricci curvature we prove discrete analogues of fundamental results by Bakry–Émery and Otto–Villani. Furthermore we show that Ricci curvature bounds are preserved under tensorisation. As a special case we obtain the sharp Ricci curvature lower bound for the discrete hypercube.

Contents

1. Introduction

In two independent contributions Sturm [41] and Lott and Villani [27] solved the long-standing open problem of defining a synthetic notion of Ricci curvature for a large class of metric measure spaces.

The key observation, proved in [39], is that on a Riemannian manifold $\mathcal{M}$, the Ricci curvature is bounded from below by some constant $\kappa \in \mathbb{R}$, if and only if the Boltzmann–Shannon entropy $\mathcal{H}(\rho) = \int \rho \log \rho \, \mathrm{dvol}$ is κ-convex along geodesics in the L^2-Wasserstein space of probability measures on $\mathcal{M}$. The latter condition does not appeal to the geometric structure of $\mathcal{M}$, but only requires a metric (to define the L^2-Wasserstein metric W_2) and a reference measure (to define the entropy $\mathcal{H}$). Therefore this condition can be used in order to define a notion of Ricci curvature lower boundedness on more

Date: July 17, 2012.
JM is supported by Rubicon subsidy 680-50-0901 of the Netherlands Organisation for Scientific Research (NWO).

general metric measure spaces. This notion turns out to be stable under Gromov–Hausdorff convergence and it implies a large number of functional inequalities with sharp constants. The theory of metric measure spaces with Ricci curvature bounds in the sense of Lott, Sturm, and Villani is still under active development [2, 3].

However, the condition of Lott–Sturm–Villani does not apply if the L^2-Wasserstein space over $\mathcal{X}$ does not contain geodesics. Unfortunately, this is the case if the underlying space is discrete (even if the underlying space consists of only two points). The aim of the present paper is to develop a variant of the theory of Lott–Sturm–Villani, which does apply to discrete spaces.

In order to circumvent the nonexistence of Wasserstein geodesics, we replace the L^2-Wasserstein metric by a different metric $\mathcal{W}$, which has been introduced in [28]. There, it has been shown that the heat flow associated with a Markov kernel on a finite set is the gradient flow of the entropy with respect to $\mathcal{W}$ (see also the independent work [15] containing related results for Fokker-Planck equations on graphs, as well as [31], where this gradient flow structure has been discovered in the setting of reaction-diffusion systems). In this sense, $\mathcal{W}$ takes over the role of the Wasserstein metric, since it is known since the seminal work by Jordan, Kinderlehrer, and Otto that the heat flow on $\mathbb{R}^n$ is the gradient flow of the entropy [23] (see [2, 18, 19, 20, 32, 36] for variations and generalisations). Convexity along $\mathcal{W}$-geodesics may thus be regarded as a discrete analogue of McCann's displacement convexity [29], which corresponds to convexity along W_2-geodesics in a continuous setting.

Since every pair of probability densities on $\mathcal{X}$ can be joined by a $\mathcal{W}$-geodesic, it is possible to define a notion of Ricci curvature in the spirit of Lott–Sturm–Villani by requiring geodesic convexity of the entropy with respect to the metric $\mathcal{W}$. This possibility has already been indicated in [28]. We shall show that this notion of Ricci curvature shares a number of properties which make the LSV definition so powerful: in particular, it is stable under tensorisation and implies a number of functional inequalities, including a modified logarithmic Sobolev inequality, and a Talagrand-type inequality involving the metric $\mathcal{W}$.

Main results. Let us now discuss the contents of this paper in more detail. We work with an irreducible Markov kernel $K : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+$ on a finite set $\mathcal{X}$, i.e., we assume that

$$\sum_{y \in \mathcal{X}} K(x, y) = 1$$

for all $x \in \mathcal{X}$, and that for every $x, y \in \mathcal{X}$ there exists a sequence $\{x_i\}_{i=0}^n \in \mathcal{X}$ such that $x_0 = x$, $x_n = y$ and $K(x_{i-1}, x_i) > 0$ for all $1 \le i \le n$. Basic Markov chain theory guarantees the existence of a unique stationary probability measure (also called steady state) π on $\mathcal{X}$, i.e.,

$$\sum_{x \in \mathcal{X}} \pi(x) = 1 \qquad \text{and} \qquad \pi(y) = \sum_{x \in \mathcal{X}} \pi(x) K(x, y)$$

for all $y \in \mathcal{X}$. We assume that π is *reversible* for K, which means that the detailed balance equations

$$K(x,y)\pi(x) = K(y,x)\pi(y) \tag{1.1}$$

hold for $x, y \in \mathcal{X}$.

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* on $\mathcal{X}$. The subset consisting of those probability densities that are strictly positive is denoted by $\mathscr{P}_*(\mathcal{X})$. We consider the metric $\mathcal{W}$ defined for $\rho_0, \rho_1 \in \mathscr{P}(\mathcal{X})$ by

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

where the infimum runs over all sufficiently regular curves $\rho : [0,1] \to \mathscr{P}(\mathcal{X})$ and $\psi : [0,1] \to \mathbb{R}^{\mathcal{X}}$ satisfying the 'continuity equation'

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t}\rho_t(x) + \displaystyle\sum_{y\in\mathcal{X}} (\psi_t(y)-\psi_t(x))\hat\rho_t(x,y)K(x,y) \;=\; 0 \qquad \forall x \in \mathcal{X} \ , \\ \rho(0) = \rho_0 \ , \qquad \rho(1) = \rho_1 \ . \end{cases} \tag{1.2}$$

Here, given $\rho \in \mathscr{P}(\mathcal{X})$, we write $\hat\rho(x,y) := \int_0^1 \rho(x)^{1-p}\rho(y)^p \,\mathrm{d}p$ for the logarithmic mean of $\rho(x)$ and $\rho(y)$. The relevance of the logarithmic mean in this setting is due to the identity

$$\rho(x) - \rho(y) = \hat\rho(x,y)(\log\rho(x) - \log\rho(y)) \ ,$$

which somewhat compensates for the lack of a 'discrete chain rule'. The definition of $\mathcal{W}$ can be regarded as a discrete analogue of the Benamou–Brenier formula [7]. Let us remark that if $t \mapsto \rho_t$ is differentiable at some t and ρ_t belongs to $\mathscr{P}_*(\mathcal{X})$, then the continuity equation (1.2) is satisfied for some $\psi_t \in \mathbb{R}^{\mathcal{X}}$, which is unique up to an additive constant (see [28, Proposition 3.26]).

Since the metric $\mathcal{W}$ is Riemannian in the interior $\mathscr{P}_*(\mathcal{X})$, it makes sense to consider gradient flows in $(\mathscr{P}_*(\mathcal{X}), \mathcal{W})$ and it has been proved in [28] that the heat flow associated with the continuous time Markov semigroup $P_t = e^{t(K-I)}$ is the gradient flow of the entropy

$$\mathcal{H}(\rho) = \sum_{x\in\mathcal{X}} \pi(x)\rho(x)\log\rho(x) \ , \tag{1.3}$$

with respect to the Riemannian structure determined by $\mathcal{W}$.

In this paper we shall show that every pair of densities $\rho_0, \rho_1 \in \mathscr{P}(\mathcal{X})$ can be joined by a constant speed geodesic. Therefore the following definition in the spirit of Lott–Sturm–Villani seems natural.

Definition 1.1. *We say that K has* non-local 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*

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

In this case, we shall use the notation

$$\mathrm{Ric}(K) \geq \kappa \ .$$

Remark 1.2. Instead of requiring convexity along all geodesics it will be shown to be equivalent to require that every pair of densities $\rho_0, \rho_1 \in \mathscr{P}(\mathcal{X})$ can be joined by a constant speed geodesic along which the entropy is κ-convex. Another equivalent condition would be to impose a lower bound on the Hessian of $\mathcal{H}$ in the interior $\mathscr{P}_*(\mathcal{X})$ (see Theorem 4.5 below for the details).

One of the main contributions of this paper is a tensorisation result for non-local Ricci curvature, which we will now describe. For $1 \leq i \leq n$, let K_i be an irreducible and reversible Markov kernel on a finite set $\mathcal{X}_i$, and let π_i denote the corresponding invariant probability measure. Let $K_{(i)}$ denote the lift of K_i to the product space $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_n$, defined for $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$ by

$$K_{(i)}(\mathbf{x}, \mathbf{y}) = \begin{cases} K_i(x_i, y_i), & \text{if } x_j = y_j \text{ for all } j \neq i, \\ 0, & \text{otherwise.} \end{cases}$$

For a sequence $\{\alpha_i\}_{1 \leq i \leq n}$ of nonnegative numbers with $\sum_{i=1}^n \alpha_i = 1$, we consider the weighted product chain, determined by the kernel

$$K_\alpha := \sum_{i=1}^n \alpha_i K_{(i)} \ .$$

Its reversible probability measure is the product measure $\pi = \pi_1 \otimes \cdots \otimes \pi_n$.

Theorem 1.3 (Tensorisation of Ricci bounds)**.** *Assume that* $\mathrm{Ric}(K_i) \geq \kappa_i$ *for* $i = 1, \ldots, n$*. Then we have*

$$\mathrm{Ric}(K_\alpha) \ \geq \ \min_i \alpha_i \kappa_i \ .$$

Tensorisation results have also been obtained for other notions of Ricci curvature, including the ones by Lott–Sturm–Villani [41, Proposition 4.16] and Ollivier [34, Proposition 27]. In both cases the proof does not extend to our setting, and completely different ideas are needed here.

As a consequence, we obtain a lower bound on the non-local Ricci curvature for (the kernel K_n of the simple random walk on) the discrete hypercube $\{0, 1\}^n$, which turns out to be optimal.

Corollary 1.4. *For* $n \geq 1$ *we have* $\mathrm{Ric}(K_n) \geq \frac{2}{n}$.

The hypercube is a fundamental building block for applications in mathematical physics and theoretical computer science, and the problem of proving "displacement convexity" on this space has been an open problem that motivated the recent paper by Ollivier and Villani [35], in which a Brunn–Minkowski inequality was obtained.

Another aspect that we wish to single out at this stage is the fact that Ricci bounds imply a number of functional inequalities, which are natural discrete counterparts to powerful inequalities in a continuous setting. In particular, we obtain discrete counterparts to the results by Bakry–Émery [5] and Otto–Villani [37].

To state the results we consider the Dirichlet form

$$\mathcal{E}(\varphi,\psi) \;=\; \frac{1}{2}\sum_{x,y\in\mathcal{X}} \big(\varphi(x)-\varphi(y)\big)\big(\psi(x)-\psi(y)\big)K(x,y)\pi(x)$$

defined for functions $\varphi,\psi:\mathcal{X}\to\mathbb{R}$. Furthermore, we consider the functional

$$\mathcal{I}(\rho) \;=\; \mathcal{E}(\rho,\log\rho)$$

defined for $\rho\in\mathscr{P}(\mathcal{X})$, with the convention that $\mathcal{I}(\rho)=+\infty$ if ρ does not belong to $\mathscr{P}_*(\mathcal{X})$. Its significance here is due to the fact that it is the time-derivative of the entropy along the heat flow: $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}(P_t\rho)=-\mathcal{I}(P_t\rho)$. In this sense, $\mathcal{I}$ can be regarded as a discrete version of the Fisher information.

Theorem 1.5 (Functional inequalities)**.** *Let K be an irreducible and reversible Markov kernel on a finite set $\mathcal{X}$.*

(1) If $\mathrm{Ric}(K)\geq\kappa$ for some $\kappa\in\mathbb{R}$, then the HWI-inequality

$$\mathcal{H}(\rho) \;\leq\; \mathcal{W}(\rho,\mathbf{1})\sqrt{\mathcal{I}(\rho)}-\frac{\kappa}{2}\mathcal{W}(\rho,\mathbf{1})^2 \qquad (\mathrm{H}\mathcal{W}\mathrm{I}(\kappa))$$

holds for all $\rho\in\mathscr{P}(\mathcal{X})$.

(2) If $\mathrm{Ric}(K)\geq\lambda$ for some $\lambda>0$, then the modified logarithmic Sobolev inequality

$$\mathcal{H}(\rho) \;\leq\; \frac{1}{2\lambda}\mathcal{I}(\rho) \qquad (\mathrm{MLSI}(\lambda))$$

holds for all $\rho\in\mathscr{P}(\mathcal{X})$.

(3) If K satisfies (MLSI(λ)) *for some $\lambda>0$, then the modified Talagrand inequality*

$$\mathcal{W}(\rho,\mathbf{1}) \;\leq\; \sqrt{\frac{2}{\lambda}\mathcal{H}(\rho)} \qquad (\mathrm{T}_{\mathcal{W}}(\lambda))$$

holds for all $\rho\in\mathscr{P}(\mathcal{X})$.

(4) If K satisfies ($\mathrm{T}_{\mathcal{W}}(\lambda)$) *for some $\lambda>0$, then the Poincaré inequality*

$$\|\varphi\|^2_{L^2(\mathcal{X},\pi)} \;\leq\; \frac{1}{\lambda}\mathcal{E}(\varphi,\varphi) \qquad (\mathrm{P}(\lambda))$$

holds for all functions $\varphi:\mathcal{X}\to\mathbb{R}$.

Here, $\mathbf{1}$ denotes the density of the stationary measure π.

The first inequality in Theorem 1.5 is a discrete counterpart to the HWI-inequality from Otto and Villani [37], with the difference that the metric W_2 has been replaced by $\mathcal{W}$.

The second result is as a discrete version of the celebrated criterion by Bakry-Émery [5], who proved the corresponding result on Riemannian manifolds. Classically, the Bakry-Émery criterion applies to weighted Riemannian manifolds $(\mathcal{M},e^{-V}\mathrm{vol}_{\mathcal{M}})$, and asks for a lower bound on the generalised Ricci curvature given by $\mathrm{Ric}_{\mathcal{M}}+\mathrm{Hess}\,V$. As in our setting we allow for general K and π, the potential V is already incorporated in K and π, and our notion of Ricci curvature can be thought of as the analogue of this generalised Ricci curvature.

The modified logarithmic Sobolev inequality (MLSI) is motivated by the fact that it yields an explicit rate of exponential decay of the entropy along the heat flow. It has been extensively studied (see, for example, [11, 14]), along with different discrete logarithmic Sobolev inequalities in the literature (for example, [4, 10]).

The third part is a discrete counterpart to a famous result by Otto and Villani [37], who showed that the logarithmic Sobolev inequality implies the so-called T_2-inequality; recall that the T_p-inequality is the analogue of $\mathrm{T}_{\mathcal{W}}$, in which $\mathcal{W}$ is replaced by W_p, for $1 \le p < \infty$. These inequalities have been extensively studied in recent years. We refer to [21] for a survey and to [40] for a study of the T_1-inequality in a discrete setting.

The modified Talagrand inequality $\mathrm{T}_{\mathcal{W}}$ that we consider is new. This inequality combines some of the good properties of T_1 and T_2, as we shall now discuss.

Like T_1, it is weak enough to be applicable in a discrete setting. In fact, we shall prove that $\mathrm{T}_{\mathcal{W}}(\lambda)$ holds on the discrete hypercube $\{0,1\}^n$ with the optimal constant $\lambda = \frac{2}{n}$. By contrast, the T_2-inequality does not even hold on the two-point space, and it has been an open problem to find an adequate substitute.

Like T_2, and unlike T_1, $\mathrm{T}_{\mathcal{W}}$ is strong enough to capture spectral information. In fact, the fourth part in Theorem 1.5 asserts that it implies a Poincaré inequality with constant λ.

Furthermore, we shall show that $\mathrm{T}_{\mathcal{W}}$ yields good bounds on the sub-Gaussian constant, in the sense that

$$\mathbb{E}_\pi\Big[e^{t(\varphi - \mathbb{E}_\pi[\varphi])}\Big] \le \exp\left(\frac{t^2}{4\lambda}\right) \tag{1.4}$$

for all $t > 0$ and all functions $\varphi : \mathcal{X} \to \mathbb{R}$ that are Lipschitz constant 1 with respect to the graph norm. Here, we use the notation $\mathbb{E}_\pi[\varphi] = \sum_{x\in\mathcal{X}} \varphi(x)\pi(x)$. As is well known, this estimate yields the concentration inequality

$$\pi\big(\varphi - \mathbb{E}_\pi[\varphi] \ge h\big) \le e^{-\lambda h^2}$$

for all $h > 0$. The proof of (1.4) relies on the fact, proved in Section 2, that the metric $\mathcal{W}$ can be bounded from below by W_1 (with respect to the graph metric), so that $\mathrm{T}_{\mathcal{W}}(\lambda)$ implies a $T_1(2\lambda)$-inequality, which is known to be equivalent to the sub-Gaussian inequality [8].

The proof of Theorem 1.5 follows the approach by Otto and Villani. On a technical level, the proofs are simpler in the discrete case, since heuristic arguments from Otto and Villani are essentially rigorous proofs in our setting, and no additional PDE arguments are required as in [37].

To summarise, we have the following sequence of implications, for any $\lambda > 0$:

$$\mathrm{Ric}(K) \ge \lambda \;\Rightarrow\; \mathrm{MLSI}(\lambda) \;\Rightarrow\; \mathrm{T}_{\mathcal{W}}(\lambda) \;\Rightarrow\; \begin{cases} \mathrm{P}(\lambda) \\ \mathrm{T}_1(2\lambda) \end{cases}.$$

Other notions of Ricci curvature. This is of course not the first time that a notion of Ricci curvature has been introduced for discrete spaces, but the notion considered here appears to be the closest in spirit to the one by Lott–Sturm–Villani. Furthermore, it seems to be the first that yields natural analogues of the results by Bakry–Émery and Otto–Villani.

A different notion of Ricci curvature has been introduced by Ollivier [33, 34]. This notion is also based on ideas from optimal transport, and uses the L^1-Wasserstein metric W_1, which behaves better in a discrete setting than W_2. Ollivier's criterion has the advantage of being easy to check in many examples. Furthermore, in some interesting cases it yields functional inequalities with good – yet non-optimal – constants. Moreover, Ollivier does not assume reversibility, whereas this is strongly used in our approach. It is not completely clear how Ollivier's notion relates to the one by Lott–Sturm–Villani (see [35] for a discussion). Furthermore, it does not seem to be directly comparable to the concept studied here, as it relies on a metric on the underlying space, which is not the case in our approach.

In the setting of graphs, Ollivier's Ricci curvature has been further studied in the recent preprints [6, 22, 24].

Another approach has been taken by Lin and Yau [26], who defined Ricci curvature in terms of the heat semigroup.

Bonciocat and Sturm [12] followed a different approach to modify the Lott–Sturm–Villani criterion, in which they circumvented the lack of midpoints in the L_2-Wasserstein metric by allowing for approximate midpoints. A Brunn–Minkowski inequality in this spirit has been proved on the discrete hypercube by Ollivier and Villani [35].

Organisation of the paper. In Section 2 we collect basic properties of the metric $\mathcal{W}$ and formulate an equivalent definition that is more convenient to work with in some situations. Geodesics in the $\mathcal{W}$-metric are studied in Section 3. In particular, it is shown that every pair of densities can be joined by a constant speed geodesic. In Section 4 we present the definition of non-local Ricci curvature and give a characterisation in terms of the Hessian of the entropy. Section 5 contains a criterion that allows us to give lower bounds on the Ricci curvature in some basic examples, including the discrete circle and the discrete hypercube. A tensorisation result is contained in Section 6. Finally, we introduce new versions of well-known functional inequalities in Section 7 and prove implications between these and known inequalities.

Note added. After essentially finishing this paper, the authors have been informed about the preprint [30], in which geodesic convexity of the entropy for Markov chains has been studied, as well. The results obtained in that paper and this one do not overlap significantly and have been obtained independently.

Acknowledgement. The authors are grateful to Nicola Gigli and Karl-Theodor Sturm for stimulating discussions on this paper and related topics. They thank the anonymous referees for detailed comments and valuable suggestions.

2. The metric $\mathcal{W}$

In this section we shall study some basic properties of the metric $\mathcal{W}$. Throughout we shall work with an irreducible and reversible Markov kernel K on a finite set $\mathcal{X}$. The unique steady state will be denoted by π, and we shall write $P_t := e^{t(K-I)}$, $t \geq 0$, to denote the corresponding Markov semigroup.

We start by introducing some notation.

2.1. **Notation.** For $\varphi \in \mathbb{R}^{\mathcal{X}}$ 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

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

With this notation we have

$$\Delta := \nabla\cdot\nabla = K - I \ ,$$

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 $\Phi,\Psi \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$,

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

From now on we shall fix a function $\theta : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ satisfying the following assumptions:

Assumption 2.1. *The function θ has the following properties:*

(A1) *(Regularity): θ is continuous on $\mathbb{R}_+\times\mathbb{R}_+$ and C^∞ on $(0,\infty)\times(0,\infty)$;*
(A2) *(Symmetry): $\theta(s,t) = \theta(t,s)$ for $s,t \geq 0$;*
(A3) *(Positivity, normalisation): $\theta(s,t) > 0$ for $s,t > 0$ and $\theta(1,1) = 1$;*
(A4) *(Zero at the boundary): $\theta(0,t) = 0$ for all $t \geq 0$;*
(A5) *(Monotonicity): $\theta(r,t) \leq \theta(s,t)$ for all $0 \leq r \leq s$ and $t \geq 0$;*
(A6) *(Positive homogeneity): $\theta(\lambda s,\lambda t) = \lambda\theta(s,t)$ for $\lambda > 0$ and $s,t \geq 0$;*
(A7) *(Concavity): the function $\theta : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is concave.*

It is easily checked that these assumptions imply that θ is bounded from above by the arithmetic mean:

$$\theta(s,t) \ \leq \ \frac{s+t}{2} \qquad \forall s,t \geq 0 \ . \tag{2.1}$$

In the next result we collect some properties of the function θ, which turn out be very useful in obtaining non-local Ricci curvature bounds.

Lemma 2.2. *For all $s,t,u,v > 0$ we have*

$$s\cdot\partial_1\theta(s,t) + t\cdot\partial_2\theta(s,t) \ = \ \theta(s,t) \ , \tag{2.2}$$
$$s\cdot\partial_1\theta(u,v) + t\cdot\partial_2\theta(u,v) - \theta(s,t) \ \geq \ 0 \ . \tag{2.3}$$

Proof. The equality (2.2) follows immediately from the homogeneity (A6) by noting that the left-hand side equals $\frac{\mathrm{d}}{\mathrm{d}r}\big|_{r=1}\theta(rs,rt)$. Let us prove (2.3). Note that by the concavity (A7) of θ the gradient $\nabla\theta$ is a monotone operator from $\mathbb{R}^2_+$ to $\mathbb{R}^2$. Hence, for all $s,t,x,y > 0$ we have

$$(s-x)\Big(\partial_1\theta(s,t) - \partial_1\theta(x,y)\Big) + (t-y)\Big(\partial_2\theta(s,t) - \partial_2\theta(x,y)\Big) \ \leq \ 0 \ .$$

By the homogeneity (A6) both $\partial_1\theta$ and $\partial_2\theta$ are 0-homogeneous. Taking now, in particular $x = \varepsilon u, y = \varepsilon v$ and letting $\varepsilon \to 0$ we obtain

$$s\Big(\partial_1\theta(s,t) - \partial_1\theta(u,v)\Big) + t\Big(\partial_2\theta(s,t) - \partial_2\theta(u,v)\Big) \ \leq\ 0\ .$$

From this we deduce (2.3) by an application of (2.2). □

The most important example for our purposes is the logarithmic mean defined by

$$\theta(s,t) := \int_0^1 s^{1-p} t^p \,\mathrm{d}p = \frac{s-t}{\log s - \log t}\ ,$$

the latter expression being valid if $s,t > 0$ and $s \neq t$. For $\rho \in \mathscr{P}(\mathcal{X})$ and $x, y \in \mathcal{X}$ we define

$$\hat\rho(x,y) = \theta(\rho(x), \rho(y))\ .$$

For a fixed $\rho \in \mathscr{P}(\mathcal{X})$ it will be useful to consider the Hilbert space $\mathcal{G}_\rho$ consisting of all (equivalence classes of) functions $\Psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$, endowed with the inner product

$$\langle \Phi, \Psi\rangle_\rho := \frac{1}{2}\sum_{x,y\in\mathcal{X}} \Phi(x,y)\Psi(x,y)\hat\rho(x,y)K(x,y)\pi(x)\ . \tag{2.4}$$

Here we identify functions that coincide on the set $\{(x,y) \in \mathcal{X}\times\mathcal{X} : \hat\rho(x,y)K(x,y) > 0\}$. The operator ∇ can then be considered as a linear operator $\nabla : L^2(\mathcal{X}) \to \mathcal{G}_\rho$, whose negative adjoint is the *ρ-divergence operator* $(\nabla_\rho\cdot) : \mathcal{G}_\rho \to L^2(\mathcal{X})$ given by

$$(\nabla_\rho \cdot \Psi)(x) := \frac{1}{2}\sum_{y\in\mathcal{X}} (\Psi(x,y) - \Psi(y,x))\hat\rho(x,y)K(x,y)\ .$$

2.2. **Equivalent Definitions of the Metric $\mathcal{W}$.** We shall now state the definition of the metric $\mathcal{W}$ as defined in [28]. Here and in the rest of the paper we will use the shorthand notation

$$\mathcal{A}(\rho,\psi) := \|\nabla\psi\|_\rho^2 = \frac{1}{2}\sum_{x,y\in\mathcal{X}} (\psi(y)-\psi(x))^2\hat\rho(x,y)K(x,y)\pi(x)$$

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

Definition 2.3. *For $\bar\rho_0, \bar\rho_1 \in \mathscr{P}(\mathcal{X})$ we define*

$$\mathcal{W}(\bar\rho_0,\bar\rho_1)^2 := \inf\left\{ \int_0^1 \mathcal{A}(\rho_t,\psi_t)\,\mathrm{d}t \ :\ (\rho,\psi)\in\mathcal{CE}_1(\bar\rho_0,\bar\rho_1)\right\}\ ,$$

where for $T > 0$, $\mathcal{CE}_T(\bar\rho_0,\bar\rho_1)$ denotes the collection of pairs (ρ,ψ) satisfying the following conditions:

$$\begin{cases} (i) & \rho : [0,T] \to \mathbb{R}^{\mathcal{X}} \ \textit{is}\ C^\infty\ ;\\ (ii) & \rho_0 = \bar\rho_0\ , \qquad \rho_T = \bar\rho_1\ ;\\ (iii) & \rho_t \in \mathscr{P}(\mathcal{X}) \ \textit{for all}\ t\in[0,T]\ ;\\ (iv) & \psi : [0,T]\to\mathbb{R}^{\mathcal{X}} \ \textit{is measurable}\ ;\\ (v) & \textit{For all}\ x\in\mathcal{X}\ \textit{and all}\ t\in(0,T)\ \textit{we have}\\ & \dot\rho_t(x) + \displaystyle\sum_{y\in\mathcal{X}} \big(\psi_t(y)-\psi_t(x)\big)\hat\rho_t(x,y)K(x,y) = 0\ . \end{cases} \tag{2.5}$$

Using the notation introduced above, the continuity equation in (v) can be written as

$$\dot{\rho}_t + \nabla \cdot (\hat{\rho} \nabla \psi) = 0 \,. \tag{2.6}$$

Definition 2.3 is the same as the one in [28], except that slightly different regularity conditions have been imposed on ρ. We shall shortly see that both definitions are equivalent.

The following results on the metric $\mathcal{W}$ have been proved in [28].

Theorem 2.4. *The following assertions hold.*

(1) *The space $(\mathscr{P}(\mathcal{X}), \mathcal{W})$ is a complete metric space, compatible with the Euclidean topology.*
(2) *The restriction of $\mathcal{W}$ to $\mathscr{P}_*(\mathcal{X})$ is the Riemannian distance induced by the following Riemannian structure:*
   - *the tangent space of $\rho \in \mathscr{P}_*(\mathcal{X})$ can be identified with the set*
   $$T_\rho := \{ \nabla \psi \ : \ \psi \in \mathbb{R}^{\mathcal{X}} \}$$
   by means of the following identification: given a smooth curve $(-\varepsilon, \varepsilon) \ni t \mapsto \rho_t \in \mathscr{P}_*(\mathcal{X})$ with $\rho_0 = \rho$, there exists a unique element $\nabla \psi_0 \in T_\rho$, such that the continuity equation (2.5)*(v) holds at $t = 0$.*
   - *The Riemannian metric on T_ρ is given by the inner product*
$$\langle \nabla \varphi, \nabla \psi \rangle_\rho = \frac{1}{2} \sum_{x,y \in \mathcal{X}} (\varphi(x) - \varphi(y))(\psi(x) - \psi(y)) \hat{\rho}(x,y) K(x,y) \pi(x) \,.$$
(3) *If θ is the logarithmic mean, i.e., $\theta(s,t) = \int_0^1 s^{1-p} t^p \, \mathrm{d}p$, then the heat flow is the gradient flow of the entropy, in the sense that for any $\rho \in \mathscr{P}(\mathcal{X})$ and $t > 0$, we have $\rho_t := P_t \rho \in \mathscr{P}_*(\mathcal{X})$ and*
$$D_t \rho_t = - \operatorname{grad} \mathcal{H}(\rho_t) \,. \tag{2.7}$$

Remark 2.5. If ρ belongs to $\mathscr{P}_*(\mathcal{X})$, then the gradient flow equation (2.7) also holds for $t = 0$.

Remark 2.6. The relevance of the logarithmic mean can be seen as follows. The heat equation $\dot{\rho}_t = \Delta \rho_t = \nabla \cdot (\nabla \rho_t)$ can be rewritten as a continuity equation (2.6) provided that

$$\nabla \psi = - \frac{\nabla \rho}{\hat{\rho}} \,.$$

On the other hand, an easy computation (see [28, Proposition 4.2 and Corollary 4.3]) shows that under the identification above, the gradient of the entropy is given by

$$\operatorname{grad}_{\mathcal{W}} \mathcal{H}(\rho) = \nabla \log \rho \,.$$

Combining these observations, we infer that the heat flow is the gradient flow of the entropy with respect to $\mathcal{W}$, precisely when

$$\frac{\nabla \rho}{\hat{\rho}} = \nabla \log \rho \,,$$

that is, when θ is the logarithmic mean.

This argument shows that the same heat flow can also be identified as the gradient flow of the functional $\mathcal{F}(\rho) = \sum_{x\in\mathcal{X}} f(\rho(x))\pi(x)$ for any smooth function $f : \mathbb{R} \to \mathbb{R}$ with $f'' > 0$, if one replaces the logarithmic mean by $\theta(r, s) = \frac{r-s}{f'(r)-f'(s)}$. We refer to [28] for the details.

Our next aim is to provide an equivalent formulation of the definition of $\mathcal{W}$, which may seem less intuitive at first sight, but offers several technical advantages. First, the continuity equation becomes linear in V and ρ, which allows us to exploit the concavity of θ. Second, this formulation is more stable so that we can prove existence of minimizers in the class $\mathcal{CE}_0'(\bar\rho_0, \bar\rho_1)$. Similar ideas have already been developed in a continuous setting in [17], where a general class of transportation metrics was constructed based on the usual continuity equation in $\mathbb{R}^n$.

An important role will be played by the function $\alpha : \mathbb{R}\times\mathbb{R}^2_+ \to \mathbb{R}\cup\{+\infty\}$ defined by

$$\alpha(x, s, t) = \begin{cases} 0\ , & \theta(s,t) = 0 \text{ and } x = 0\ ,\\ \frac{x^2}{\theta(s,t)}\ , & \theta(s,t) \neq 0\ ,\\ +\infty\ , & \theta(s,t) = 0 \text{ and } x \neq 0\ . \end{cases}$$

The following observation will be useful.

Lemma 2.7. *The function α is lower semicontinuous and convex.*

Proof. This is easily checked using (A7) and the convexity of the function $(x, y) \mapsto \frac{x^2}{y}$ on $\mathbb{R} \times (0,\infty)$. □

Given $\rho \in \mathscr{P}(\mathcal{X})$ and $V \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ we define

$$\mathcal{A}'(\rho, V) := \frac{1}{2}\sum_{x,y\in\mathcal{X}} \alpha(V(x,y), \rho(x), \rho(y))K(x,y)\pi(x)\ ,$$

and we set

$$\mathcal{CE}_T'(\bar\rho_0, \bar\rho_1) := \{(\rho,\psi)\ :\ (i'), (ii), (iii), (iv'), (v') \text{ hold}\}\ ,$$

where

$$\begin{cases} (i') & \rho : [0,T] \to \mathbb{R}^{\mathcal{X}} \text{ is continuous ;}\\ (iv') & V : [0,T] \to \mathbb{R}^{\mathcal{X}\times\mathcal{X}} \text{ is locally integrable ;}\\ (v') & \text{For all } x\in\mathcal{X} \text{ we have in the sense of distributions}\\ & \dot\rho_t(x) + \frac{1}{2}\sum_{y\in\mathcal{X}} \big(V_t(x,y) - V_t(y,x)\big)K(x,y) = 0\ . \end{cases} \tag{2.8}$$

The continuity equation in (v') can equivalently be written as

$$\dot\rho_t + \nabla\cdot V = 0\ .$$

As an immediate consequence of Lemma 2.7 we obtain the following convexity of $\mathcal{A}'$.

Corollary 2.8. *Let $\rho^i \in \mathscr{P}(\mathcal{X})$ and $V^i \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ for $i = 0, 1$. For $\tau \in [0, 1]$ set $\rho^\tau := (1-\tau)\rho^0 + \tau\rho^1$ and $V^\tau := (1-\tau)V^0 + \tau V^1$. Then we have*

$$\mathcal{A}'(\rho^\tau, V^\tau)\ \le\ (1-\tau)\mathcal{A}'(\rho^0, V^0) + \tau\mathcal{A}'(\rho^1, V^1)\ .$$

Now we have the following reformulation of Definition 2.3.

Lemma 2.9. *For $\bar\rho_0, \bar\rho_1 \in \mathscr{P}(\mathcal{X})$ we have*

$$\mathcal{W}(\bar\rho_0,\bar\rho_1)^2 = \inf\left\{ \int_0^1 \mathcal{A}'(\rho_t, V_t)\,\mathrm{d}t \ : \ (\rho,V)\in \mathcal{CE}_1'(\bar\rho_0,\bar\rho_1)\right\} .$$

Furthermore, if $\bar\rho_0, \bar\rho_1 \in \mathscr{P}_*(\mathcal{X})$, condition (iv) in (2.5) can be reinforced into: "$\psi : [0,T] \to \mathbb{R}^{\mathcal{X}}$ is C^∞".

Proof. The inequality "$\geq$" follows easily by noting that the infimum is taken over a larger set. Indeed, given a pair $(\rho,\psi)\in\mathcal{CE}_1(\bar\rho_0,\bar\rho_1)$ we obtain a pair $(\rho,V)\in\mathcal{CE}_1'(\bar\rho_0,\bar\rho_1)$ by setting $V_t(x,y) = \nabla\psi_t(x,y)\hat\rho_t(x,y)$ and we have $\mathcal{A}'(\rho_t,V_t) = \mathcal{A}(\rho_t,\psi_t)$.

To show the opposite inequality "$\leq$", we fix an arbitrary pair $(\rho,V)\in\mathcal{CE}_1'(\bar\rho_0,\bar\rho_1)$. It is sufficient to show that for every $\varepsilon>0$ there exists a pair $(\rho^\varepsilon,\psi^\varepsilon)\in\mathcal{CE}_1(\bar\rho_0,\bar\rho_1)$ such that

$$\int_0^1 \mathcal{A}(\rho_t^\varepsilon,\psi_t^\varepsilon)\,\mathrm{d}t \ \leq \ \int_0^1 \mathcal{A}'(\rho_t,V_t)\,\mathrm{d}t + \varepsilon\ .$$

For this purpose we first regularise (ρ,V) by a mollification argument. We thus define $(\widetilde\rho,\widetilde V) : [-\varepsilon,1+\varepsilon]\to\mathscr{P}(\mathcal{X})\times\mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ by

$$(\widetilde\rho_t,\widetilde V_t) \ = \ \begin{cases} (\rho(0),0)\ , & t\in[-\varepsilon,\varepsilon)\ ,\\ (\rho(\frac{t-\varepsilon}{1-2\varepsilon}), \frac{1}{1-2\varepsilon}V(\frac{t-\varepsilon}{1-2\varepsilon}))\ , & t\in[\varepsilon,1-\varepsilon)\ ,\\ (\rho(1),0)\ , & t\in[1-\varepsilon,1+\varepsilon]\ , \end{cases}$$

and take a nonnegative smooth function $\eta:\mathbb{R}\to\mathbb{R}_+$ which vanishes outside of $[-\varepsilon,\varepsilon]$, is strictly positive on $(-\varepsilon,\varepsilon)$ and satisfies $\int\eta(s)\,\mathrm{d}s = 1$. For $t\in[0,1]$ we define

$$\rho_t^\varepsilon \ = \ \int \eta(s)\widetilde\rho_{t+s}\,\mathrm{d}s\ , \qquad V_t^\varepsilon \ = \ \int\eta(s)\widetilde V_{t+s}\,\mathrm{d}s\ .$$

Now $t\mapsto\rho_t^\varepsilon$ is C^∞ and using the continuity of ρ it is easy to check that $(\rho^\varepsilon,V^\varepsilon)\in\mathcal{CE}_1'(\bar\rho_0,\bar\rho_1)$. Moreover, using the convexity from Corollary 2.8 we can estimate

$$\begin{aligned}\int_0^1 \mathcal{A}'(\rho_t^\varepsilon,V_t^\varepsilon)\,\mathrm{d}t \ &\leq \ \int_0^1\int\eta(s)\mathcal{A}'(\widetilde\rho_{t+s},\widetilde V_{t+s})\,\mathrm{d}s\,\mathrm{d}t\\ &\leq \ \int_{-\varepsilon}^{1+\varepsilon}\mathcal{A}'(\widetilde\rho_t,\widetilde V_t)\,\mathrm{d}t \ = \ \frac{1}{1-2\varepsilon}\int_0^1\mathcal{A}'(\rho_t,V_t)\,\mathrm{d}t\ .\end{aligned}$$

To proceed further, we may assume without loss of generality that $V(x,y) = 0$ whenever $K(x,y)=0$. The fact that $\int_0^1\mathcal{A}'(\rho_t,V_t)\,\mathrm{d}t$ is finite implies that the set $\{t : \hat\rho_t(x,y) = 0 \text{ and } V_t(x,y)\neq 0\}$ is negligible for all $x,y\in\mathcal{X}$. Taking properties (A3) and (A4) of the function θ into account, this implies that for the convolved quantities the corresponding set $\{t : \hat\rho_t^\varepsilon(x,y) = 0 \text{ and } V_t^\varepsilon(x,y)\neq 0\}$ is empty for all $x,y\in\mathcal{X}$. Hence there exists a measurable function $\Psi^\varepsilon : [0,1]\to\mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ satisfying

$$V_t^\varepsilon(x,y) \ = \ \Psi_t^\varepsilon(x,y)\hat\rho_t^\varepsilon(x,y) \qquad \text{for all } x,y\in\mathcal{X} \text{ and all } t\in[0,1]\ . \tag{2.9}$$

It remains to find a function $\psi^\varepsilon : [0,1]\to\mathbb{R}^{\mathcal{X}}$ such that $\nabla_{\rho_t^\varepsilon}\cdot\Psi_t^\varepsilon = \nabla_{\rho_t^\varepsilon}\cdot\nabla\psi_t^\varepsilon$. Let $\mathcal{P}_\rho$ denote the orthogonal projection in $\mathcal{G}_\rho$ onto the range of ∇. Then there exists a measurable function $\psi^\varepsilon : [0,1] \to \mathbb{R}^{\mathcal{X}}$ such that $\mathcal{P}_{\rho_t^\varepsilon}\Psi_t^\varepsilon = \nabla\psi_t^\varepsilon$. The orthogonal decomposition

$$\mathcal{G}_{\rho_t^\varepsilon} \;=\; \mathrm{Ran}(\nabla) \oplus^\perp \mathrm{Ker}(\nabla_{\rho_t^\varepsilon}^*) \tag{2.10}$$

implies that $\nabla_{\rho_t^\varepsilon} \cdot \Psi_t^\varepsilon = \nabla_{\rho_t^\varepsilon} \cdot \nabla\psi_t^\varepsilon$, hence $(\rho^\varepsilon, \psi^\varepsilon) \in \mathcal{CE}_1(\bar\rho_o, \bar\rho_1)$. Using the decomposition (2.10) once more, we infer that $\langle \nabla\psi_t^\varepsilon, \nabla\psi_t^\varepsilon\rangle_{\rho_t^\varepsilon} \le \langle \Psi_t^\varepsilon, \Psi_t^\varepsilon\rangle_{\rho_t^\varepsilon}$. This implies $\mathcal{A}(\rho_t^\varepsilon, \psi_t^\varepsilon) \le \mathcal{A}'(\rho_t^\varepsilon, V_t^\varepsilon)$ and finishes the proof of the first assertion.

If $\bar\rho_0$ and $\bar\rho_1$ belong to $\mathscr{P}_*(\mathcal{X})$, one can follow the argument in [28, Lemma 3.30] and construct a curve $(\breve\rho, \breve V) \in \mathcal{CE}_1'(\bar\rho_0, \bar\rho_1)$ such that $\breve\rho_t \in \mathscr{P}_*(\mathcal{X})$ for $t \in [0,1]$ and

$$\int_0^1 \mathcal{A}'(\breve\rho_t, \breve V_t)\,\mathrm{d}t \;\le\; \int_0^1 \mathcal{A}'(\rho_t, V_t)\,\mathrm{d}t + \varepsilon\,.$$

Then one can apply the argument above. In this case, $\rho_t^\varepsilon(x) > 0$ for all $x \in \mathcal{X}$ and $t \in [0,1]$, and therefore the function $\Psi^\varepsilon : [0,1] \to \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ is C^∞. Furthermore, since the orthogonal projection P_ρ depends smoothly on $\rho \in \mathscr{P}_*(\mathcal{X})$, the function $\psi^\varepsilon : [0,1] \to \mathbb{R}^{\mathcal{X}}$ is smooth as well. □

Remark 2.10. In [28] the metric $\mathcal{W}$ has been defined as in Definition 2.3, with the difference that (i) in (2.8) was replaced by "$\rho : [0,T] \to \mathscr{P}(\mathcal{X})$ is piecewise C^1". Therefore Lemma 2.9 shows, in particular, that Definition 2.3 coincides with the original definition of $\mathcal{W}$ from [28].

2.3. **Basic properties of $\mathcal{W}$.** As an application of Lemma 2.9 we shall prove the following convexity result, which is a discrete counterpart of the well-known fact that the squared L^2-Wasserstein distance over Euclidean space is convex with respect to linear interpolation (see, for example, [17, Theorem 5.11]).

Proposition 2.11 (Convexity of the squared distance)**.** *For $i, j = 0, 1$, let $\rho_i^j \in \mathscr{P}(\mathcal{X})$, and for $\tau \in [0,1]$ set $\rho_i^\tau := (1-\tau)\rho_i^0 + \tau\rho_i^1$. Then*

$$\mathcal{W}(\rho_0^\tau, \rho_1^\tau)^2 \le (1-\tau)\mathcal{W}(\rho_0^0, \rho_1^0)^2 + \tau\mathcal{W}(\rho_0^1, \rho_1^1)^2\,.$$

Proof. Let $\varepsilon > 0$. For $j = 0, 1$ we may take a pair $(\rho^j, V^j) \in \mathcal{CE}'(\rho_0^j, \rho_1^j)$ with

$$\int_0^1 \mathcal{A}'(\rho_t^j, V_t^j)\,\mathrm{d}t \le \mathcal{W}^2(\rho_0^j, \rho_1^j) + \varepsilon$$

in view of Lemma 2.9. For $\tau \in [0,1]$ we set

$$\rho_t^\tau := (1-\tau)\rho_t^0 + \tau\rho_t^1\,, \qquad V_t^\tau := (1-\tau)V_t^0 + \tau V_t^1\,.$$

It then follows that $(\rho^\tau, V^\tau) \in \mathcal{CE}_1'(\rho_0^\tau, \rho_1^\tau)$, hence by Corollary 2.8,

$$\begin{aligned}\mathcal{W}(\rho_0^\tau, \rho_1^\tau)^2 &\le \int_0^1 \mathcal{A}'(\rho_t^\tau, V_t^\tau)\,\mathrm{d}t\\ &\le (1-\tau)\int_0^1 \mathcal{A}'(\rho_t^0, V_t^0)\,\mathrm{d}t + \tau\int_0^1 \mathcal{A}'(\rho_t^1, V_t^1)\,\mathrm{d}t\\ &= (1-\tau)\mathcal{W}(\rho_0^0, \rho_1^0)^2 + \tau\mathcal{W}(\rho_0^1, \rho_1^1)^2 + \varepsilon\,.\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, this completes the proof. □

In this section we compare $\mathcal{W}$ to some commonly used metrics. A first result of this type (see [28, Lemma 3.10]) gives a lower bound on $\mathcal{W}$ in terms of the total variation metric

$$d_{TV}(\rho_0, \rho_1) = \sum_{x \in \mathcal{X}} \pi(x)|\rho_0(x) - \rho_1(x)| \ .$$

Here, more generally, we shall compare $\mathcal{W}$ to various Wasserstein distances. Given a metric d on $\mathcal{X}$ and $1 \le p < \infty$, recall that the L^p-Wasserstein metric $W_{p,d}$ on $\mathscr{P}(\mathcal{X})$ is defined by

$$W_{p,d}(\rho_0, \rho_1) := \inf \left\{ \Big(\sum_{x,y \in \mathcal{X}} d(x,y)^p q(x,y) \Big)^{\frac{1}{p}} \ \Big| \ q \in \Gamma(\rho_0, \rho_1) \right\} , \qquad (2.11)$$

where $\Gamma(\rho_0, \rho_1)$ denotes the set of all couplings between ρ_0 and ρ_1, i.e.,

$$\Gamma(\rho_0, \rho_1) := \Big\{ q : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_+ \ \Big| \ \sum_{y \in \mathcal{X}} q(x,y) = \rho_0(x)\pi(x) \ , \quad \sum_{x \in \mathcal{X}} q(x,y) = \rho_1(y)\pi(y) \Big\} \ .$$

It is well known (see, for example, [43, Theorem 4.1]) that the infimum in (2.11) is attained; as usual we shall denote the collection of minimizers by $\Gamma_o(\rho_0, \rho_1)$.

In our setting there are various metrics on $\mathcal{X}$ that are natural to consider. In particular,

- the *graph distance* d_g with respect to the graph structure on $\mathcal{X}$ induced by K (i.e., $\{x, y\}$ is an edge iff $K(x, y) > 0$).
- the metric $d_{\mathcal{W}}$, that is, the restriction of $\mathcal{W}$ from $\mathscr{P}(\mathcal{X})$ to $\mathcal{X}$ under the identification of points in $\mathcal{X}$ with the corresponding Dirac masses:

$$d_{\mathcal{W}}(x,y) := \mathcal{W}\left(\frac{\mathbf{1}_{\{x\}}}{\pi(x)}, \frac{\mathbf{1}_{\{y\}}}{\pi(y)} \right) \ .$$

The induced L^p-Wasserstein distances will be denoted by $W_{p,g}$ and $W_{p,\mathcal{W}}$ respectively.

We shall now prove lower and upper bounds for the metric $\mathcal{W}$ in terms of suitable Wasserstein metrics. We start with the lower bounds. Let us remark that, unlike most other results in this paper, the second inequality in the following result relies on the normalisation $\sum_{y \in \mathcal{X}} K(x, y) = 1$.

Proposition 2.12 (Lower bounds for $\mathcal{W}$)**.** *For all probability densities $\rho_0, \rho_1 \in \mathscr{P}(\mathcal{X})$ we have*

$$\frac{1}{\sqrt{2}} d_{TV}(\rho_0, \rho_1) \le \sqrt{2} W_{1,g}(\rho_0, \rho_1) \le \mathcal{W}(\rho_0, \rho_1) \ . \qquad (2.12)$$

Proof. Note that $d_{tr} \le d_g$, where $d_{tr}(x,y) = \mathbf{1}_{x \neq y}$ denotes the trivial distance. Therefore, the first bound follows from the fact that d_{TV} is the L^1-Wasserstein distance induced by d_{tr} (see [42, Theorem 1.14]).

In order to prove the second bound, we fix $\varepsilon > 0$, take $\bar\rho_0, \bar\rho_1 \in \mathscr{P}(\mathcal{X})$ and $(\rho, \psi) \in \mathcal{CE}_1(\bar\rho_0, \bar\rho_1)$ with

$$\left(\int_0^1 \mathcal{A}(\rho_t, \psi_t)\, \mathrm{d}t \right)^{\frac{1}{2}} \le \mathcal{W}(\bar\rho_0, \bar\rho_1) + \varepsilon \ .$$

Using the continuity equation from (2.5) we obtain for any $\varphi : \mathcal{X} \to \mathbb{R}$,

$$\begin{aligned}
&\Big| \sum_{x\in\mathcal{X}} \varphi(x)(\rho_0(x) - \rho_1(x))\pi(x) \Big| \\
&= \Big| \int_0^1 \sum_{x\in\mathcal{X}} \varphi(x)\dot\rho_t(x)\pi(x)\, \mathrm{d}t \Big| \\
&= \Big| \int_0^1 \sum_{x,y\in\mathcal{X}} \varphi(x)\big(\psi_t(x) - \psi_t(y)\big)\hat\rho_t(x,y)K(x,y)\pi(x)\, \mathrm{d}t \Big| \\
&= \Big| \int_0^1 \langle \nabla\varphi, \nabla\psi_t \rangle_{\rho_t}\, \mathrm{d}t \Big| \\
&\le \left(\int_0^1 \|\nabla\varphi\|_{\rho_t}^2\, \mathrm{d}t \right)^{1/2} \left(\int_0^1 \|\nabla\psi_t\|_{\rho_t}^2\, \mathrm{d}t \right)^{1/2} \\
&= \left(\int_0^1 \|\nabla\varphi\|_{\rho_t}^2\, \mathrm{d}t \right)^{1/2} (\mathcal{W}(\bar\rho_0, \bar\rho_1) + \varepsilon) \ .
\end{aligned}$$

Let $[\varphi]_{\mathrm{Lip}}$ denote the Lipschitz constant of φ with respect to the graph distance d_g, i.e.,

$$[\varphi]_{\mathrm{Lip}} := \sup_{x\neq y} \frac{|\varphi(x) - \varphi(y)|}{d_g(x,y)} \ .$$

Applying the inequality (2.1) and using the fact that $d_g(x,y) = 1$ if $x \neq y$ and $K(x,y) > 0$, we infer that

$$\begin{aligned}
\|\nabla\varphi\|_{\rho_t}^2 &= \frac{1}{2} \sum_{x,y\in\mathcal{X}} \big(\varphi(x) - \varphi(y)\big)^2 K(x,y)\hat\rho_t(x,y)\pi(x) \\
&\le \frac{1}{4}[\varphi]_{\mathrm{Lip}}^2 \sum_{x,y\in\mathcal{X}} K(x,y)\big(\rho_t(x) + \rho_t(y)\big)\pi(x) \\
&= \frac{1}{2}[\varphi]_{\mathrm{Lip}}^2 \sum_{x\in\mathcal{X}} \rho_t(x)\pi(x) \sum_{y\in\mathcal{X}} K(x,y) \\
&= \frac{1}{2}[\varphi]_{\mathrm{Lip}}^2 \ .
\end{aligned}$$

The Kantorovich–Rubinstein Theorem (see, for example, [42, Theorem 1.14]) yields

$$W_{1,g}(\bar\rho_0, \bar\rho_1) = \sup_{\varphi : [\varphi]_{\mathrm{Lip}} \le 1} \Big| \sum_{x\in\mathcal{X}} \varphi(x)(\bar\rho_0(x) - \bar\rho_1(x))\pi(x) \Big| \le \frac{\mathcal{W}(\bar\rho_0, \bar\rho_1) + \varepsilon}{\sqrt{2}} \ ,$$

which completes the proof, since $\varepsilon > 0$ is arbitrary. □

Before stating the upper bounds, we provide a simple relation between d_g and $d_{\mathcal{W}}$.

Lemma 2.13. *For $x, y \in \mathcal{X}$ we have*

$$d_{\mathcal{W}}(x,y) \le \frac{c}{\sqrt{k}} d_g(x,y) \ ,$$

where

$$c = \int_{-1}^{1} \frac{\mathrm{d}r}{\sqrt{2\theta(1-r,1+r)}} < \infty \qquad \text{and} \qquad k = \min_{(x,y)\ :\ K(x,y)>0} K(x,y) \ .$$

If θ is the logarithmic mean, then $c \approx 1.56$.

Proof. Let $\{x_i\}_{i=0}^n$ be a sequence in $\mathcal{X}$ with $x_0 = x$, $x_n = y$ and $K(x_i, x_{i+1}) > 0$ for all i. We shall use the fact, proved in [28, Theorem 2.4], that the $\mathcal{W}$-distance between two Dirac measures on a two-point space $\{a, b\}$ with transition probabilities $K(a,b) = K(b,a) = p$ is equal to $\frac{c}{\sqrt{p}}$. The concavity of θ readily implies that c is finite. Furthermore, it follows from [28, Lemma 3.14] and its proof, that for any pair $x, y \in \mathcal{X}$ with $K(x,y) > 0$, one has

$$\mathcal{W}\bigg(\frac{\mathbf{1}_{\{x\}}}{\pi(x)}, \frac{\mathbf{1}_{\{y\}}}{\pi(y)}\bigg) \le c\sqrt{\frac{\max\{\pi(x),\pi(y)\}}{K(x,y)\pi(x)}} \le \frac{c}{\sqrt{k}} \ .$$

Using the triangle inequality for $\mathcal{W}$ we obtain

$$d_{\mathcal{W}}(x,y) = \mathcal{W}\bigg(\frac{\mathbf{1}_{\{x\}}}{\pi(x)}, \frac{\mathbf{1}_{\{y\}}}{\pi(y)}\bigg) \le \sum_{i=0}^{n-1} \mathcal{W}\bigg(\frac{\mathbf{1}_{\{x_i\}}}{\pi(x_i)}, \frac{\mathbf{1}_{\{x_{i+1}\}}}{\pi(x_{i+1})}\bigg) \le \frac{nc}{\sqrt{k}} \ ,$$

hence the result follows by taking the infimum over all such sequences $\{x_i\}_{i=0}^n$. □

Now we turn to upper bounds for $\mathcal{W}$ in terms of L^2-Wasserstein distances.

Proposition 2.14 (Upper bounds for $\mathcal{W}$)**.** *For all probability densities $\rho_0, \rho_1 \in \mathscr{P}(\mathcal{X})$ we have*

$$\mathcal{W}(\rho_0,\rho_1) \le W_{2,\mathcal{W}}(\rho_0,\rho_1) \le \frac{c}{\sqrt{k}} W_{2,g}(\rho_0,\rho_1) \ , \tag{2.13}$$

where c and k are as in Lemma 2.13.

Proof. We shall prove the first bound, the second one being an immediate consequence of Lemma 2.13. For this purpose, we fix $\bar\rho_0, \bar\rho_1 \in \mathscr{P}(\mathcal{X})$ and take $q \in \Gamma_o(\bar\rho_0, \bar\rho_1)$. For all $u, v \in \mathcal{X}$, take a curve $(\rho^{u,v}, V^{u,v}) \in \mathcal{CE}'\big(\frac{\mathbf{1}_{\{u\}}}{\pi(u)}, \frac{\mathbf{1}_{\{v\}}}{\pi(v)}\big)$ with

$$\int_0^1 \mathcal{A}'(\rho_t^{u,v}, V_t^{u,v})\,\mathrm{d}t \le d_{\mathcal{W}}(u,v)^2 + \varepsilon \ ,$$

and consider the convex combination of these curves, weighted according to the optimal plan q, i.e.,

$$\rho_t := \sum_{u,v\in\mathcal{X}} q(u,v)\rho_t^{u,v} \ , \qquad V_t := \sum_{u,v\in\mathcal{X}} q(u,v) V_t^{u,v} \ .$$

It then follows that the resulting curve (ρ, V) belongs to $\mathcal{CE}'_1(\bar\rho_0, \bar\rho_1)$. Using the convexity result from Lemma 2.7 we infer that

$$\begin{aligned}\mathcal{W}(\bar\rho_0,\bar\rho_1)^2 \le \int_0^1 \mathcal{A}'(\rho_t, V_t)\,\mathrm{d}t &\le \sum_{u,v\in\mathcal{X}} q(u,v)\int_0^1 \mathcal{A}'(\rho_t^{u,v}, V_t^{u,v})\,\mathrm{d}t\\ &\le \sum_{u,v\in\mathcal{X}} q(u,v)(d_{\mathcal{W}}(u,v)^2+\varepsilon)\\ &= W_{2,\mathcal{W}}(\bar\rho_0,\bar\rho_1)^2+\varepsilon\ .\end{aligned}$$

which implies the result. □

3. Geodesics

In this section we show that the metric space $(\mathscr{P}(\mathcal{X}), \mathcal{W})$ is a geodesic space, in the sense that any two densities $\rho_0, \rho_1 \in \mathscr{P}(\mathcal{X})$ can be connected by a *(constant speed) geodesic*, that is, a curve $\gamma : [0,1] \to \mathscr{P}(\mathcal{X})$ satisfying

$$\mathcal{W}(\gamma_s,\gamma_t) = |s-t|\mathcal{W}(\gamma_0,\gamma_1)$$

for all $0 \le s, t \le 1$.

Let us first give an equivalent characterisation of the infimum in Lemma 2.9, which is invariant under reparametrisation.

Lemma 3.1. *For any $T > 0$ and $\bar\rho_0, \bar\rho_1 \in \mathscr{P}(\mathcal{X})$ we have*

$$\mathcal{W}(\bar\rho_0,\bar\rho_1) \ = \ \inf\left\{\int_0^T \sqrt{\mathcal{A}'(\rho_t,V_t)}dt \ : \ (\rho,V)\in\mathcal{CE}'_T(\bar\rho_0,\bar\rho_1)\right\}\ . \qquad (3.1)$$

Proof. Taking Lemma 2.9 into account, this follows from a standard reparametrisation argument. See [1, Lemma 1.1.4] or [17, Theorem 5.4] for details in similar situations. □

Theorem 3.2. *For all $\bar\rho_0, \bar\rho_1 \in \mathscr{P}(\mathcal{X})$ the infimum in Lemma 2.9 is attained by a pair $(\rho, V) \in \mathcal{CE}'_1(\bar\rho_0, \bar\rho_1)$ satisfying $\mathcal{A}'(\rho_t, V_t) = \mathcal{W}(\bar\rho_0, \bar\rho_1)^2$ for a.e. $t \in [0, 1]$. In particular, the curve $(\rho_t)_{t\in[0,1]}$ is a constant speed geodesic.*

Proof. We will show existence of a minimizing curve by a direct argument. Let $(\rho^n, V^n) \in \mathcal{CE}'_1(\bar\rho_0, \bar\rho_1)$ be a minimizing sequence. Thus we can assume that

$$\sup_n \int_0^1 \mathcal{A}'(\rho_t^n, V_t^n)\,\mathrm{d}t \ < \ C$$

for some finite constant C. Without loss of generality we assume that $V_t^n(x, y) = 0$ when $K(x, y) = 0$. For $x, y \in \mathcal{X}$, define the sequence of signed Borel measures $\nu^n_{x,y}$ on $[0, 1]$ by $\nu^n_{x,y}(\mathrm{d}t) := V_t^n(x, y)\ \mathrm{d}t$. For every Borel set $B \subset [0, 1]$ we can give the following bound on the total variation of these measures:

$$\|\nu^n_{x,y}\|(B) \ \le \ \int_B |V_t^n(x,y)|\,\mathrm{d}t \le \ \sqrt{C'}\int_B \sqrt{\alpha(V_t^n(x,y),\rho_t^n(x),\rho_t^n(y))}\,\mathrm{d}t\ ,$$

where we used the fact that $\rho(x) \le \max\{\pi(z)^{-1} : z \in \mathcal{X}\} =: C' < \infty$ for $\rho \in \mathscr{P}(\mathcal{X})$. Using Hölder's inequality we obtain

$$\begin{aligned} \sum_{x,y\in\mathcal{X}} \|\nu^n_{x,y}\|(B)K(x,y)\pi(x) \ &\le\ \sqrt{2C'\,\mathrm{Leb}(B)}\left(\int_0^1 \mathcal{A}'(\rho^n_t,V^n_t)\,\mathrm{d}t\right)^{\frac12} \\ &\le\ \sqrt{2CC'\,\mathrm{Leb}(B)}\ . \end{aligned} \tag{3.2}$$

In particular, the total variation of the measures $\nu^n_{x,y}$ is bounded uniformly in n. Hence we can extract a subsequence (still indexed by n) such that for all $x,y \in \mathcal{X}$ the measures $\nu^n_{x,y}$ converge weakly* to some finite signed Borel measure $\nu_{x,y}$. The estimate (3.2) also shows that $\nu_{x,y}$ is absolutely continuous with respect to the Lebesgue measure. Thus there exists $V : [0,1] \to \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ such that $\nu_{x,y}(\mathrm{d}t) := V_t(x,y)\,\mathrm{d}t$. We claim that, along the same subsequence, ρ^n converges pointwise to a function $\rho : [0,1] \to \mathscr{P}(\mathcal{X})$. Indeed, using the continuity of $t \mapsto \rho^n_t$ one derives from the continuity equation (v') in (2.8) that for $s \in [0,1]$ and every $x \in \mathcal{X}$,

$$\rho^n_s - \rho^n_0 \ =\ \frac12 \int_0^s \sum_{y\in\mathcal{X}} (V^n_t(y,x) - V^n_t(x,y))K(x,y)\,\mathrm{d}t\ . \tag{3.3}$$

The weak* convergence of $\nu^n_{x,y}$ implies (see [1, Prop. 5.1.10]) the convergence of the right-hand side of (3.3). Since $\rho^n_0 = \bar\rho_0$ for all n, this yields the desired convergence of ρ^n_s for all s, and one easily checks that $(\rho,V) \in \mathcal{CE}'_1(\rho_0,\rho_1)$. The weak* convergence of $\nu^n_{x,y}$ further implies that the measures $\rho^n_t(x)\mathrm{d}t$ converge weakly* to $\rho_t(x)\mathrm{d}t$. Applying a general result on the lower-semicontinuity of integral functionals (see [13, Thm. 3.4.3]) and taking into account Lemma 2.7, we obtain

$$\int_0^1 \mathcal{A}'(\rho_t,V_t)\,\mathrm{d}t\ \le\ \liminf_n \int_0^1 \mathcal{A}'(\rho^n_t,V^n_t)\,\mathrm{d}t\ =\ \mathcal{W}(\bar\rho_0,\bar\rho_1)^2\ .$$

Hence the pair (ρ,V) is a minimizer of the variational problem in the definition of $\mathcal{W}$. Finally, Lemma 3.1 yields

$$\int_0^1 \sqrt{\mathcal{A}'(\rho_t,V_t)}\,\mathrm{d}t\ \ge\ \mathcal{W}(\bar\rho_0,\bar\rho_1)\ =\ \left(\int_0^1 \mathcal{A}'(\rho_t,V_t)\,\mathrm{d}t\right)^{\frac12}\ ,$$

which implies that $\mathcal{A}'(\rho_t,V_t) = W(\bar\rho_0,\bar\rho_1)^2$ for a.e. $t \in [0,1]$.

The fact that $(\rho_t)_t$ is a constant speed geodesic follows now by another application of Lemma 3.1. □

We shall now give a characterisation of absolutely continuous curves in the metric space $(\mathscr{P}(\mathcal{X}),\mathcal{W})$ and relate their length to their minimal action. First we recall some notions from the theory of analysis in metric spaces. A curve $(\rho_t)_{t\in[0,T]}$ in $\mathscr{P}(\mathcal{X})$ is called *absolutely continuous w.r.t.* $\mathcal{W}$ if there exists $m \in L^1(0,T)$ such that

$$\mathcal{W}(\rho_s,\rho_t)\ \le\ \int_s^t m(r)\,\mathrm{d}r \quad \text{for all } 0 \le s \le t \le T\ .$$

If (ρ_t) is absolutely continuous, then its *metric derivative*

$$|\rho_t'| := \lim_{h\to 0} \frac{\mathcal{W}(\rho_{t+h}, \rho_t)}{|h|}$$

exists for a.e. $t \in [0,T]$ and satisfies $|\rho_t'| \le m(t)$ a.e. (see [1, Theorem 1.1.2]).

Proposition 3.3 (Metric velocity)**.** *A curve $(\rho_t)_{t\in[0,T]}$ is absolutely continuous with respect to $\mathcal{W}$ if and only if there exists a measurable function $V : [0,T] \to \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ such that $(\rho, V) \in \mathcal{CE}'_T(\rho_0, \rho_T)$ and*

$$\int_0^T \sqrt{\mathcal{A}'(\rho_t, V_t)}dt < \infty .$$

In this case we have $|\rho_t'|^2 \le \mathcal{A}'(\rho_t, V_t)$ for a.e. $t \in [0,T]$ and there exists an aalmost everywhere uniquely defined function $\widetilde{V} : [0,1] \to \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ such that $(\rho, \widetilde{V}) \in \mathcal{CE}'_T(\rho_0, \rho_T)$ and $|\rho_t'|^2 = \mathcal{A}'(\rho_t, \widetilde{V}_t)$ for a.e. $t \in [0,T]$.

Proof. The proof follows from the very same arguments as in [17, Thm. 5.17]. To construct the velocity field $\widetilde{V}$, the curve ρ is approximated by curves (ρ^n, V^n) which are piecewise minimizing. The velocity field $\widetilde{V}$ is then defined as a subsequential limit of the velocity fields V^n. In our case, existence of this limit is guaranteed by a compactness argument similar to the one in the proof of Theorem 3.2. □

For later use we state an explicit formula for the geodesic equations in $\mathscr{P}_*(\mathcal{X})$ from [28, Proposition 3.4]. Since the interior $\mathscr{P}_*(\mathcal{X})$ of $\mathscr{P}(\mathcal{X})$ is Riemannian by Theorem 2.4, local existence and uniqueness of geodesics is guaranteed by standard Riemannian geometry.

Proposition 3.4. *Let $\bar\rho \in \mathscr{P}_*(\mathcal{X})$ and $\bar\psi \in \mathbb{R}^{\mathcal{X}}$. On a sufficiently small time interval around 0, the unique constant speed geodesic with $\rho_0 = \bar\rho$ and initial tangent vector $\nabla\psi_0 = \nabla\bar\psi$ satisfies the following equations:*

$$\begin{cases} \partial_t\rho_t(x) + \sum_{y\in\mathcal{X}} (\psi_t(y) - \psi_t(x))\hat\rho_t(x,y)K(x,y) = 0 , \\ \partial_t\psi_t(x) + \frac{1}{2}\sum_{y\in\mathcal{X}} \big(\psi_t(x) - \psi_t(y)\big)^2 \partial_1\theta(\rho_t(x), \rho_t(y))K(x,y) = 0 . \end{cases} \tag{3.4}$$

4. Ricci curvature

In this section we initiate the study of a notion of Ricci curvature lower boundedness in the spirit of Lott, Sturm, and Villani [27, 41]. Furthermore, we present a characterisation, which we shall use to prove Ricci bounds in concrete examples.

As before, we fix an irreducible and reversible Markov kernel K on a finite set $\mathcal{X}$ with steady state π. The associated Markov semigroup shall be denoted by $(P_t)_{t\ge 0}$.

Assumption 4.1. *Throughout the remainder of the paper we assume that θ is the logarithmic mean.*

We are now ready to state the definition, which has already been given in [28, Definition 1.3].

Definition 4.2. *We say that K has* non-local Ricci curvature bounded from below by $\kappa \in \mathbb{R}$ *and write* $\mathrm{Ric}(K) \geq \kappa$, *if the following holds: for every constant speed geodesic* $(\rho_t)_{t\in[0,1]}$ *in* $(\mathscr{P}(\mathcal{X}), \mathcal{W})$ *we have*

$$\mathcal{H}(\rho_t) \leq (1-t)\mathcal{H}(\rho_0) + t\mathcal{H}(\rho_1) - \frac{\kappa}{2} t(1-t)\mathcal{W}(\rho_0,\rho_1)^2 \;. \tag{4.1}$$

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

$$\begin{aligned}
\mathcal{B}(\rho,\psi) :=&\ \frac{1}{2}\big\langle \hat\Delta\rho \cdot \nabla\psi, \nabla\psi\big\rangle_\pi - \big\langle \hat\rho \cdot \nabla\psi \ , \nabla\Delta\psi\big\rangle_\pi \\
=&\ \frac{1}{4}\sum_{x,y,z\in\mathcal{X}} \big(\psi(x)-\psi(y)\big)^2 \Big(\partial_1\theta\big(\rho(x),\rho(y)\big)\big(\rho(z)-\rho(x)\big)K(x,z) \\
&\qquad + \partial_2\theta\big(\rho(x),\rho(y)\big)\big(\rho(z)-\rho(y)\big)K(y,z)\Big)K(x,y)\pi(x) \\
&- \frac{1}{2}\sum_{x,y,z\in\mathcal{X}} \Big(K(x,z)\big(\psi(z)-\psi(x)\big) - K(y,z)\big(\psi(z)-\psi(y)\big)\Big) \\
&\qquad \times \big(\psi(x)-\psi(y)\big)\hat\rho(x,y)K(x,y)\pi(x) \ ,
\end{aligned} \tag{4.2}$$

where

$$\hat\Delta\rho(x,y) := \partial_1\theta(\rho(x),\rho(y))\Delta\rho(x) + \partial_2\theta(\rho(x),\rho(y))\Delta\rho(y) \;.$$

The significance of $\mathcal{B}(\rho,\psi)$ is mainly due to the following result:

Proposition 4.3. *For* $\rho \in \mathscr{P}_*(\mathcal{X})$ *and* $\psi \in \mathbb{R}^{\mathcal{X}}$ *we have*

$$\big\langle \mathrm{Hess}\,\mathcal{H}(\rho)\nabla\psi \ , \ \nabla\psi\big\rangle_\rho = \mathcal{B}(\rho,\psi) \;.$$

Proof. Take (ρ,ψ) satisfying the geodesic equations (3.4), so that

$$\big\langle \mathrm{Hess}\,\mathcal{H}(\rho_t)\nabla\psi_t \ , \ \nabla\psi_t\big\rangle_{\rho_t} = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{H}(\rho_t) \;.$$

Using the continuity equation we obtain

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{H}(\rho_t) &= -\big\langle 1+\log\rho_t, \nabla\cdot(\hat\rho_t\nabla\psi_t)\big\rangle_\pi \\
&= \big\langle \nabla\log\rho_t, \hat\rho_t\cdot\nabla\psi_t\big\rangle_\pi \\
&= \big\langle \nabla\rho_t, \nabla\psi_t\big\rangle_\pi \;.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{H}(\rho_t) &= \big\langle \nabla\partial_t\rho_t, \nabla\psi_t\big\rangle_\pi + \big\langle \nabla\rho_t, \nabla\partial_t\psi_t\big\rangle_\pi \\
&= -\big\langle \partial_t\rho_t, \Delta\psi_t\big\rangle_\pi - \big\langle \Delta\rho_t, \partial_t\psi_t\big\rangle_\pi \;.
\end{aligned}$$

Using the continuity equation we obtain

$$\begin{aligned}
\big\langle \partial_t\rho_t, \Delta\psi_t\big\rangle_\pi &= -\big\langle \nabla\cdot(\hat\rho\nabla\psi_t), \Delta\psi_t\big\rangle_\pi \\
&= \big\langle \hat\rho_t\nabla\psi_t, \nabla\Delta\psi_t\big\rangle_\pi = \big\langle \nabla\psi_t, \nabla\Delta\psi_t\big\rangle_{\rho_t} \;.
\end{aligned}$$

Furthermore, applying the geodesic equations (3.4) and the detailed balance equations (1.1), we infer that

$$\begin{aligned}
&\big\langle \Delta\rho_t, \partial_t\psi_t\big\rangle_\pi \\
&= -\frac{1}{2}\sum_{x,y,z\in\mathcal{X}} \big(\psi_t(x)-\psi_t(y)\big)^2 \partial_1\theta\big(\rho_t(x),\rho_t(y)\big) \\
&\qquad\qquad \times \big(\rho_t(z)-\rho_t(x)\big)K(x,y)K(x,z)\pi(x) \\
&= -\frac{1}{4}\sum_{x,y,z\in\mathcal{X}} \big(\psi_t(x)-\psi_t(y)\big)^2 \Big(\partial_1\theta\big(\rho_t(x),\rho_t(y)\big)\big(\rho_t(z)-\rho_t(x)\big)K(x,z) \\
&\qquad\qquad + \partial_2\theta\big(\rho_t(x),\rho_t(y)\big)\big(\rho_t(z)-\rho_t(y)\big)K(y,z)\Big)K(x,y)\pi(x) \\
&= -\frac{1}{2}\big\langle \hat\Delta\rho_t \cdot \nabla\psi_t, \nabla\psi_t\big\rangle_\pi\ .
\end{aligned}$$

Combining the latter three identities, we arrive at

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{H}(\rho_t) = -\big\langle \nabla\psi_t, \nabla\Delta\psi_t\big\rangle_{\rho_t} + \frac{1}{2}\big\langle \hat\Delta\rho_t \cdot \nabla\psi_t, \nabla\psi_t\big\rangle_\pi\ ,$$

which is the desired identity. □

Our next aim is to show that κ-convexity of $\mathcal{H}$ along geodesics is equivalent to a lower bound of the Hessian of $\mathcal{H}$ in $\mathscr{P}_*(\mathcal{X})$. Since the Riemannian metric on $(\mathscr{P}(\mathcal{X}), \mathcal{W})$ degenerates at the boundary, this is not an obvious result. In particular, in order to prove the implication "(4) ⇒ (3)" below we cannot directly apply the equivalence between the so-called EVI (4.4) and the usual gradient flow equation, which holds on complete Riemannian manifolds (see, for example, [43, Proposition 23.1]). Therefore, we take a different approach, based on an argument by Daneri and Savaré [16], which avoids delicate regularity issues for geodesics. An additional benefit of this approach is that we expect it to apply in a more general setting where the underlying space $\mathcal{X}$ is infinite, and finite-dimensional Riemannian techniques do not apply at all.

Remark 4.4. The quantity $\mathcal{B}(\rho,\psi)$ arises naturally in the Eulerian approach to the Wasserstein metric, as developed in [16, 38]. In fact, in a crucial argument from [16], the authors consider a certain two-parameter family of measures (ρ_t^s) and functions (ψ_t^s) on a Riemannian manifold $\mathcal{M}$, and show that

$$\partial_s\mathcal{H}(\rho_t^s) + \frac{1}{2}\partial_t\int_{\mathcal{M}} |\nabla\psi_t^s|^2 \,\mathrm{d}\rho_t^s = -B(\rho_t^s,\psi_t^s)\ , \tag{4.3}$$

where

$$B(\rho,\psi) := \int_{\mathcal{M}} \Big(\frac{1}{2}\Delta(|\nabla\psi|^2) - \langle\nabla\psi,\nabla\Delta\psi\rangle\Big)\,\mathrm{d}\rho\ .$$

Since Bochner's formula asserts that

$$B(\rho,\psi) := \int_{\mathcal{M}} |D^2\psi|^2 + \mathrm{Ric}(\nabla\psi,\nabla\psi)\,\mathrm{d}\rho\ ,$$

one obtains a lower bound on B if the Ricci curvature is bounded from below. The lower bound on B can be used to prove an evolution variational inequality, which in turn yields convexity of the entropy along W_2-geodesics.

In our setting, the quantity $\mathcal{B}(\rho,\psi)$ can be regarded as a discrete analogue of $B(\rho,\psi)$. Therefore the inequality $\mathcal{B}(\rho,\psi) \geq \kappa\mathcal{A}(\rho,\psi)$ could be interpreted as a one-sided Bochner inequality, which allows us to adapt the strategy from [16] to the discrete setting.

In the following result and the rest of the paper we shall use the notation

$$\frac{\mathrm{d}^+}{\mathrm{d}t} f(t) = \limsup_{h\downarrow 0} \frac{f(t+h)-f(t)}{h} \ .$$

Theorem 4.5. *Let $\kappa \in \mathbb{R}$. For an irreducible and reversible Markov kernel $(\mathcal{X}, K)$ the following assertions are equivalent:*

(1) $\mathrm{Ric}(K) \geq \kappa$ *;*
(2) For all $\rho,\nu \in \mathscr{P}(\mathcal{X})$, the following 'evolution variational inequality' holds for all $t \geq 0$:

$$\frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}t}\mathcal{W}^2(P_t\rho,\nu) + \frac{\kappa}{2}\mathcal{W}^2(P_t\rho,\nu) \leq \mathcal{H}(\nu) - \mathcal{H}(P_t\rho) \ ; \tag{4.4}$$

(3) For all $\rho,\nu \in \mathscr{P}_*(\mathcal{X})$, (4.4) holds for all $t \geq 0$;
(4) For all $\rho \in \mathscr{P}_*(\mathcal{X})$ and $\psi \in \mathbb{R}^{\mathcal{X}}$ we have

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

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

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

(6) For all $\bar\rho_0, \bar\rho_1 \in \mathscr{P}_*(\mathcal{X})$ there exists a constant speed geodesic $(\rho_t)_{t\in[0,1]}$ satisfying $\rho_0 = \bar\rho_0$, $\rho_1 = \bar\rho_1$, and (4.1).

Proof. "(3) ⇒ (2)": This is a special case of [16, Theorem 3.3].

"(2) ⇒ (1)": This follows by applying [16, Theorem 3.2] to the metric space $(\mathscr{P}(\mathcal{X}),\mathcal{W})$ and the functional $\mathcal{H}$.

"(1) ⇒ (6)": This is clear in view of Theorem 3.2.

"(6) ⇒ (5)": Take $\rho \in \mathscr{P}_*(\mathcal{X})$ and $\psi \in \mathbb{R}^{\mathcal{X}}$ and consider the unique solution $(\rho_t,\psi_t)_{t\in(-\varepsilon,\varepsilon)}$ to the geodesic equations with $\rho_0 = \rho$ and $\psi_0 = \psi$ on a sufficiently small time interval around 0. Using the local uniqueness of geodesics and (6), we infer that

$$\mathrm{Hess}\,\mathcal{H}(\rho)(\nabla\psi) = \frac{\mathrm{d}^2}{\mathrm{d}t^2}\Big|_{t=0}\mathcal{H}(\rho_t) \geq \kappa\|\nabla\psi\|_\rho^2$$

(see, for example, the implication "$(ii) \Leftrightarrow (i)$" in [43, Proposition 16.2]).

"(5) ⇒ (4)": This follows from Proposition 4.3.

"(4) ⇒ (3)": We follow [16]. In view of Lemma 2.9 we can find a smooth curve $(\rho^\cdot,\psi^\cdot) \in \mathcal{CE}_1(\nu,\rho)$ satisfying

$$\int_0^1 \mathcal{A}(\rho^s,\psi^s)\,\mathrm{d}s < \mathcal{W}(\rho,\nu)^2 + \varepsilon \ . \tag{4.5}$$

Note in particular that $s \mapsto \rho^s$ and $s \mapsto \psi^s$ are sufficiently regular to apply Lemma 4.6 below. Using the notation from this lemma, we infer that

$$\frac{1}{2}\partial_t\mathcal{A}(\rho_t^s,\psi_t^s) + \partial_s\mathcal{H}(\rho_t^s) = -s\mathcal{B}(\rho_t^s,\psi_t^s) \ .$$

Using the assumption that $\mathcal{B} \geq \kappa\mathcal{A}$ we infer that

$$\frac{1}{2}\partial_t\Big(e^{2\kappa st}\mathcal{A}(\rho_t^s,\psi_t^s)\Big) + \partial_s\Big(e^{2\kappa st}\mathcal{H}(\rho_t^s)\Big) \leq 2\kappa t e^{2\kappa st}\mathcal{H}(\rho_t^s)\ .$$

Integration with respect to $t \in [0,h]$ and $s \in [0,1]$ yields

$$\begin{aligned}\frac{1}{2}\int_0^1 &\Big(e^{2\kappa sh}\mathcal{A}(\rho_h^s,\psi_h^s) - \mathcal{A}(\rho_0^s,\psi_0^s)\Big)\,\mathrm{d}s\\ &+ \int_0^h \Big(e^{2\kappa t}\mathcal{H}(\rho_t^1) - \mathcal{H}(\rho_t^0)\Big)\,\mathrm{d}t \leq 2\kappa\int_0^1\int_0^h t e^{2\kappa st}\mathcal{H}(\rho_t^s)\,\mathrm{d}t\,\mathrm{d}s\ .\end{aligned}$$

Arguing as in [16, Lemma 5.1] we infer that

$$\int_0^1 e^{2\kappa sh}\mathcal{A}(\rho_h^s,\psi_h^s)\,\mathrm{d}s \geq m(\kappa h)\mathcal{W}^2(P_h\rho,\nu)\ ,$$

where $m(\kappa) = \frac{\kappa e^\kappa}{\sinh(\kappa)}$. Using (4.5) together with the fact that the entropy decreases along the heat flow, we infer that

$$\begin{aligned}\frac{m(\kappa h)}{2}&\mathcal{W}^2(P_h\rho,\nu) - \frac{1}{2}\mathcal{W}^2(\rho,\nu) - \varepsilon\\ &+ E_\kappa(h)\mathcal{H}(P_h\rho) - h\mathcal{H}(\nu) \leq 2\kappa\int_0^1\int_0^h t e^{2\kappa st}\mathcal{H}(\rho_t^s)\,\mathrm{d}t\,\mathrm{d}s\ ,\end{aligned}\tag{4.6}$$

where $E_\kappa(h) := \int_0^h e^{2\kappa t}\,dt$. Since $\mathcal{H}$ is bounded, it follows that

$$\lim_{h\downarrow 0}\frac{1}{h}\int_0^1\int_0^h t e^{2\kappa st}\mathcal{H}(\rho_t^s)\,\mathrm{d}t\,\mathrm{d}s = 0\ .$$

Furthermore,

$$\lim_{h\downarrow 0}\frac{1}{h}\Big(E_\kappa(h)\mathcal{H}(P_h\rho) - h\mathcal{H}(\nu)\Big) = \mathcal{H}(\rho) - \mathcal{H}(\nu)\ .$$

Since $\varepsilon > 0$ is arbitrary, (4.6) implies that

$$\frac{\mathrm{d}^+}{\mathrm{d}h}\Big|_{h=0}\left(\frac{m(\kappa h)}{2}\mathcal{W}^2(P_h\rho,\nu)\right) + \mathcal{H}(\rho) - \mathcal{H}(\nu) \leq 0\ .$$

Taking into account that

$$\frac{\mathrm{d}^+}{\mathrm{d}h}\Big|_{h=0}\left(\frac{m(\kappa h)}{2}\mathcal{W}^2(P_h\rho,\nu)\right) = \frac{\kappa}{2}\mathcal{W}^2(\rho,\nu) + \frac{1}{2}\frac{\mathrm{d}^+}{\mathrm{d}h}\Big|_{h=0}\mathcal{W}^2(P_h\rho,\nu)\ ,$$

we obtain (4.4) for $t = 0$, which clearly implies (4.4) for all $t \geq 0$. □

The following result, which is used in the proof of Theorem 4.5, is a discrete analogue of (4.3) and the proof proceeds along the lines of [16, Lemma 4.3]. Since the details are slightly different in the discrete setting, we present a proof for the convenience of the reader.

Lemma 4.6. *Let $\{\rho^s\}_{s\in[0,1]}$ be a smooth curve in $\mathscr{P}(\mathcal{X})$. For each $t \geq 0$, set $\rho_t^s := e^{st\Delta}\rho^s$, and let $\{\psi_t^s\}_{s\in[0,1]}$ be a smooth curve in $\mathbb{R}^\mathcal{X}$ satisfying the continuity equation*

$$\partial_s\rho_t^s + \nabla\cdot(\hat{\rho}_t^s\cdot\nabla\psi_t^s) = 0\ ,\qquad s\in[0,1]\ .$$

Then the identity

$$\frac{1}{2}\partial_t \mathcal{A}(\rho_t^s, \psi_t^s) + \partial_s \mathcal{H}(\rho_t^s) = -s\mathcal{B}(\rho_t^s, \psi_t^s)$$

holds for every $s \in [0,1]$ and $t \geq 0$.

Proof. First of all, we have

$$\begin{aligned}
\partial_s \mathcal{H}(\rho_t^s) &= \big\langle 1 + \log \rho_t^s \ , \ \partial_s \rho_t^s \big\rangle_\pi \\
&= -\big\langle 1 + \log \rho_t^s \ , \ \nabla \cdot (\hat\rho_t^s \cdot \nabla \psi) \big\rangle_\pi \\
&= \big\langle \nabla \log \rho_t^s \ , \ \hat\rho_t^s \cdot \nabla \psi_t^s \big\rangle_\pi \\
&= \big\langle \nabla \rho_t^s \ , \ \nabla \psi_t^s \big\rangle_\pi \\
&= -\big\langle \psi_t^s \ , \ \Delta \rho_t^s \big\rangle_\pi \ .
\end{aligned} \tag{4.7}$$

Furthermore,

$$\begin{aligned}
\frac{1}{2}\partial_t \mathcal{A}(\rho_t^s, \psi_t^s) &= \big\langle \hat\rho_t^s \cdot \partial_t \nabla \psi_t^s \ , \ \nabla \psi_t^s \big\rangle_\pi + \frac{1}{2}\big\langle \partial_t \hat\rho_t^s \cdot \nabla \psi_t^s \ , \ \nabla \psi_t^s \big\rangle_\pi \\
&=: I_1 + I_2 \ .
\end{aligned}$$

In order to simplify I_1 we claim that

$$-\nabla \cdot \big((\partial_t \hat\rho_t^s) \cdot \nabla \psi_t^s\big) - \nabla \cdot \big(\hat\rho_t^s \cdot \partial_t \nabla \psi_t^s\big) = \Delta \rho_t^s - s\Delta\big(\nabla \cdot (\hat\rho_t^s \cdot \nabla \psi_t^s)\big) \ , \tag{4.8}$$

$$\partial_t \hat\rho_t^s = s\hat\Delta \rho_t^s \ . \tag{4.9}$$

To show (4.8), note that the left-hand side equals $\partial_t \partial_s \rho_t^s$, while the right-hand side equals $\partial_s \partial_t \rho_t^s$. The identity (4.9) follows from a straightforward calculation.

Integrating by parts repeatedly and using (4.7), (4.8) and (4.9), we obtain

$$\begin{aligned}
I_1 &= -\big\langle \psi_t^s \ , \ \nabla \cdot (\hat\rho_t^s \cdot \partial_t \nabla \psi_t^s) \big\rangle_\pi \\
&= \big\langle \psi_t^s \ , \ \Delta \rho_t^s \big\rangle_\pi - s\big\langle \psi_t^s \ , \ \Delta\big(\nabla \cdot (\hat\rho_t^s \nabla \psi_t^s)\big) \big\rangle_\pi + \big\langle \psi_t^s \ , \ \nabla \cdot \big((\partial_t \hat\rho_t^s) \nabla \psi_t^s\big) \big\rangle_\pi \\
&= -\partial_s \mathcal{H}(\rho_t^s) + s\big\langle \hat\rho_t^s \cdot \nabla \psi_t^s \ , \ \nabla \Delta \psi_t^s \big\rangle_\pi - s\big\langle \hat\Delta \rho_t^s \cdot \nabla \psi_t^s \ , \ \nabla \psi_t^s \big\rangle_\pi \ .
\end{aligned}$$

Taking into account that

$$I_2 = \frac{s}{2}\big\langle \hat\Delta \rho_t^s \cdot \nabla \psi_t^s \ , \ \nabla \psi_t^s \big\rangle_\pi \ ,$$

the result follows by summing the expressions for I_1 and I_2. □

The evolution variational inequality (4.4) has been extensively studied in the theory of gradient flows in metric spaces [1]. It readily implies a number of interesting properties for the associated gradient flow (see, for example, [16, Section 3]). Among them we single out the following κ-contractivity property.

Proposition 4.7 (κ-Contractivity of the heat flow)**.** *Let $(\mathcal{X}, K)$ be an irreducible and reversible Markov kernel satisfying* $\mathrm{Ric}(K) \geq \kappa$ *for some $\kappa \in \mathbb{R}$. Then the associated continuous time Markov semigroup $(P_t)_{t\geq 0}$ satisfies*

$$\mathcal{W}(P_t\rho, P_t\sigma) \leq e^{-\kappa t}\mathcal{W}(\rho, \sigma)$$

for all $\rho, \sigma \in \mathscr{P}(\mathcal{X})$ and $t \geq 0$.

Proof. This follows by applying [16, Proposition 3.1] to the functional $\mathcal{H}$ on the metric space $(\mathscr{P}(\mathcal{X}), \mathcal{W})$. □

5. Examples

In this section we give explicit lower bounds on the non-local Ricci curvature in several examples. Moreover, we present a simple criterion (see Proposition 5.4) for proving non-local Ricci curvature bounds. Although the assumptions seem restrictive, the criterion allows us to obtain the sharp Ricci bound for the discrete hypercube. Moreover, it can be combined with the tensorisation result from Section 6 in order to prove Ricci bounds in other nontrivial situations. To get started let us consider a particularly simple example.

Example 5.1 (The complete graph). Let $\mathcal{K}^n$ denote the complete graph on n vertices and let K_n be the simple random walk on $\mathcal{K}^n$ given by the transition kernel $K(x,y) = \frac{1}{n}$ for all $x, y \in \mathcal{K}^n$. Note that in this case π is the uniform measure. We will show that $\mathrm{Ric}(K_n) \geq \frac{1}{2} + \frac{1}{2n}$. In view of Theorem 4.5 we have to show $\mathcal{B}(\rho,\psi) \geq (\frac{1}{2} + \frac{1}{2n})\mathcal{A}(\rho,\psi)$ for all $\rho \in \mathscr{P}_*(\mathcal{X})$ and $\psi \in \mathbb{R}^{\mathcal{X}}$. Recall the definition (4.2) of the quantity $\mathcal{B}$. We calculate explicitly:

$$\begin{aligned}\big\langle \hat\rho \cdot \nabla\psi \ , \ \nabla\Delta\psi \big\rangle_\pi &= \frac{1}{2}\frac{1}{n^3}\sum_{x,y,z\in\mathcal{X}} \hat\rho(x,y)\nabla\psi(y,x)\Big[\nabla\psi(x,z) - \nabla\psi(y,z)\Big]\\ &= -\frac{1}{2}\frac{1}{n^2}\sum_{x,y}\hat\rho(x,y)\big(\nabla\psi(x,y)\big)^2 \ = \ -\mathcal{A}(\rho,\psi)\ .\end{aligned}$$

With the notation $\hat\rho_i(x,y) = \partial_i\theta(\rho(x),\rho(y))$ and using equation (2.2) we obtain further

$$\begin{aligned}\big\langle \hat\Delta\rho \ \cdot \ \nabla\psi, \nabla\psi \big\rangle_\pi &= \frac{1}{2}\frac{1}{n^3}\sum_{x,y,z}\big(\nabla\psi(x,y)\big)^2\Big[\hat\rho_1(x,y)(\rho(z)-\rho(x))\\ &\qquad\qquad + \hat\rho_2(x,y)(\rho(z)-\rho(y))\Big]\\ &= -\mathcal{A}(\rho,\psi) + \frac{1}{2}\frac{1}{n^3}\sum_{x,y,z}\big(\nabla\psi(x,y)\big)^2\Big[\hat\rho_1(x,y)\rho(z)\\ &\qquad\qquad + \hat\rho_2(x,y)\rho(z)\Big]\ .\end{aligned}$$

Keeping only the terms with $z = x$ (resp. $z = y$) in the last sum and using (2.2) again, we see

$$\big\langle \hat\Delta\rho \ \cdot \ \nabla\psi, \nabla\psi \big\rangle_\pi \ \geq \ \left(\frac{1}{n} - 1\right)\mathcal{A}(\rho,\psi)\ .$$

Summing up, we obtain $\mathcal{B} \geq (\frac{1}{2}(\frac{1}{n} - 1) + 1)\mathcal{A}$, which yields the claim.

For the rest of this section we let K be an irreducible and reversible Markov kernel on a finite set $\mathcal{X}$. In order to state the criterion and to perform calculations, it will be convenient to write a Markov chain in terms of allowed moves rather than jumps from point to point.

Let G be a set of maps from $\mathcal{X}$ to itself (the allowed moves) and consider a function $c : \mathcal{X} \times G \to \mathbb{R}_+$ (representing the jump rates).

Definition 5.2. *We call the pair* (G, c) *a* mapping representation *of* K *if the following properties hold:*

(1) The generator $\Delta = K - I$ *can be written in the form*

$$\Delta\psi(x) \;=\; \sum_{\delta\in G} \nabla_\delta\psi(x)c(x,\delta)\ , \tag{5.1}$$

where

$$\nabla_\delta\psi(x) = \psi(\delta x) - \psi(x)\ .$$

(2) For every $\delta \in G$ *there exists a unique* $\delta^{-1} \in G$ *satisfying* $\delta^{-1}(\delta(x)) = x$ *for all* x *with* $c(x,\delta) > 0$.

(3) For every $F : \mathcal{X} \times G \to \mathbb{R}$ *we have*

$$\sum_{x\in\mathcal{X},\delta\in G} F(x,\delta)c(x,\delta)\pi(x) \;=\; \sum_{x\in\mathcal{X},\delta\in G} F(\delta x,\delta^{-1})c(x,\delta)\pi(x)\ . \tag{5.2}$$

Remark 5.3. This definition is close in spirit to the recent work [14], where Γ_2-type calculations have been performed in order to prove strict convexity of the entropy along the heat flow in a discrete setting. Here, we essentially compute the second derivatives of the entropy along $\mathcal{W}$-geodesics. Since the geodesic equations are more complicated than the heat equation, the expressions that we need to work with are somewhat more involved.

Every irreducible, reversible Markov chain has a mapping representation. In fact, an explicit mapping representation can be obtained as follows. For $x, y \in \mathcal{X}$ consider the bijection $t_{\{x,y\}} : \mathcal{X} \to \mathcal{X}$ that interchanges x and y and keeps all other points fixed. Then let G be the set of all these "transpositions" and set $c(x, t_{\{x,y\}}) = K(x, y)$ and $c(x, t_{\{y,z\}}) = 0$ for $x \notin \{y, z\}$. Then (G, c) defines a mapping representation. However, in examples it is often more natural to work with a different mapping representation involving a smaller set G, as we shall see below.

It will be useful to formulate the expressions for $\mathcal{A}$ and $\mathcal{B}$ in this formalism. For this purpose, we note that (5.1) implies that

$$\sum_{y\in\mathcal{X}} F(x,y)K(x,y) \;=\; \sum_{\delta\in G} F(x,\delta x)c(x,\delta)$$

for any $F : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ vanishing on the diagonal. As a consequence we obtain

$$\mathcal{A}(\rho,\psi) \;=\; \frac{1}{2}\sum_{x\in\mathcal{X},\delta\in G} \big(\nabla_\delta\psi(x)\big)^2\hat\rho(x,\delta x)c(x,\delta)\pi(x) \tag{5.3}$$

and

$$\begin{aligned}\langle\hat\rho\nabla\psi,\nabla\Delta\psi\rangle_\pi \;=\; \frac{1}{2}\sum_{x\in\mathcal{X}}\sum_{\delta,\eta\in G} \nabla_\delta\psi(x)\Big[&\nabla_\eta\psi(\delta x)c(\delta x,\eta) \\ &- \nabla_\eta\psi(x)c(x,\eta)\Big]\hat\rho(x,\delta x)c(x,\delta)\pi(x)\ .\end{aligned} \tag{5.4}$$

Setting for convenience $\partial_i\theta(\rho(x),\rho(y)) =: \hat\rho_i(x,y)$ for $i=1,2$ we further get

$$\begin{aligned}\frac{1}{2}\langle\hat\Delta\rho\nabla\psi,\nabla\psi\rangle_\pi \;&=\; \frac{1}{4}\sum_{x,\delta,\eta}\big(\nabla_\delta\psi(x)\big)^2\Big[\hat\rho_1(x,\delta x)\nabla_\eta\rho(x)c(x,\eta)\\ &\qquad + \hat\rho_2(x,\delta x)\nabla_\eta\rho(\delta x)c(\delta x,\eta)\Big]c(x,\delta)\pi(x)\;.\end{aligned} \tag{5.5}$$

Now the expression for $\mathcal{B}(\rho,\psi)$ is obtained as the difference between the two preceding two expressions.

We are now ready to state the announced criterion, which shall be used in Examples 5.6 and 5.7 below. Intuitively, condition (ii) expresses a certain 'spatial homogeneity', saying that the jump rate in a given direction is the same before and after another jump.

Proposition 5.4. *Let K be an irreducible and reversible Markov kernel on a finite set $\mathcal{X}$ and let (G,c) be a mapping representation. Consider the following conditions:*

(i) $\delta\circ\eta \;=\; \eta\circ\delta$ *, for all* $\delta,\eta\in G$*,*
(ii) $c(\delta x,\eta) \;=\; c(x,\eta)$ *, for all* $x\in\mathcal{X},\;\delta,\eta\in G$*,*
(iii) $\delta\circ\delta = id$ *, for all* $\delta\in G$*.*

If (i) and (ii) are satisfied, then $\mathrm{Ric}(K)\geq 0$*. If moreover (iii) is satisfied, then* $\mathrm{Ric}(K)\geq 2C$*, where*

$$C := \min\{c(x,\delta) : x\in\mathcal{X}, \delta\in G \text{ such that } c(x,\delta)>0\}\;.$$

Remark 5.5. Note that requiring *(i)* and *(iii)* simultaneously imposes a very strong restriction on the graph associated with K. We prefer to state the result in this form in order to give a unified proof which applies both to the discrete circle and the discrete hypercube, with optimal constant in the latter case.

Proof of Proposition 5.4. In view of Theorem 4.5 it suffices to show that $\mathcal{B}(\rho,\psi)\geq 0$ resp. $\mathcal{B}(\rho,\psi)\geq 2C\mathcal{A}(\rho,\psi)$ for all $\rho\in\mathscr{P}_*(\mathcal{X})$ and $\psi\in\mathbb{R}^{\mathcal{X}}$. First recall that

$$\mathcal{B}(\rho,\psi) \;=\; -\langle\hat\rho\nabla\psi,\nabla\Delta\psi\rangle_\pi + \frac{1}{2}\langle\hat\Delta\rho\nabla\psi,\nabla\psi\rangle_\pi \;=:\; T_1+T_2\;.$$

Using (5.4) and conditions (i) and (ii) we can write the first summand as

$$\begin{aligned}T_1 \;&=\; -\frac{1}{2}\sum_{x,\delta,\eta}\nabla_\delta\psi(x)\Big[\nabla_\eta\psi(\delta x)-\nabla_\eta\psi(x)\Big]\hat\rho(x,\delta x)c(x,\delta)c(x,\eta)\pi(x)\\ &=\; -\frac{1}{2}\sum_{x,\delta,\eta}\nabla_\delta\psi(x)\Big[\nabla_\delta\psi(\eta x)-\nabla_\delta\psi(x)\Big]\hat\rho(x,\delta x)c(x,\delta)c(x,\eta)\pi(x)\;.\end{aligned}$$

In a similar way we shall write the second summand. Starting from (5.5) and invoking (ii) and equation (2.2) from Lemma 2.2, we obtain

$$\begin{aligned} T_2 &= \frac{1}{4}\sum_{x,\delta,\eta} \big(\nabla_\delta\psi(x)\big)^2\Big[\hat\rho_1(x,\delta x)\nabla_\eta\rho(x) \\ &\qquad\qquad + \hat\rho_2(x,\delta x)\nabla_\eta\rho(\delta x)\Big]c(x,\delta)c(x,\eta)\pi(x) \\ &= \frac{1}{4}\sum_{x,\delta,\eta} \big(\nabla_\delta\psi(x)\big)^2\Big[\hat\rho_1(x,\delta x)\rho(\eta x) \\ &\qquad\qquad + \hat\rho_2(x,\delta x)\rho(\eta\delta x) - \hat\rho(x,\delta x)\Big]c(x,\delta)c(x,\eta)\pi(x)\ . \end{aligned}$$

Using the reversibility of K in the form of (5.2), and again condition (ii) we can write

$$\begin{aligned} T_2 &= \frac{1}{4}\sum_{x,\delta,\eta} \Bigg(\big(\nabla_\delta\psi(\eta x)\big)^2\Big[\hat\rho_1(\eta x,\delta\eta x)\rho(x) + \hat\rho_2(\eta x,\delta\eta x)\rho(\delta x)\Big] \\ &\qquad - \big(\nabla_\delta\psi(x)\big)^2\hat\rho(x,\delta x)\Bigg)c(x,\delta)c(x,\eta)\pi(x)\ . \end{aligned}$$

Adding zero we obtain

$$\begin{aligned} T_2 &= \frac{1}{4}\sum_{x,\delta,\eta} \Big(\big(\nabla_\delta\psi(\eta x)\big)^2 - \big(\nabla_\delta\psi(x)\big)^2\Big)\,\hat\rho(x,\delta x)c(x,\delta)c(x,\eta)\pi(x) \\ &\quad + \frac{1}{4}\sum_{x,\delta,\eta} \big(\nabla_\delta\psi(\eta x)\big)^2\Big[\hat\rho_1(\eta x,\delta\eta x)\rho(x) + \hat\rho_2(\eta x,\delta\eta x)\rho(\delta x) \\ &\qquad\qquad - \hat\rho(x,\delta x)\Big]c(x,\delta)c(x,\eta)\pi(x) \\ &=: T_3 + T_4\ . \end{aligned}$$

Invoking the inequality (2.3) from Lemma 2.2, we immediately see that $T_4 \geq 0$. Hence we get

$$\begin{aligned} \mathcal{B}(\rho,\psi) &\geq T_1 + T_3 \\ &= \frac{1}{4}\sum_{x,\delta,\eta} \big(\nabla_\delta\psi(\eta x) - \nabla_\delta\psi(x)\big)^2\hat\rho(x,\delta x)c(x,\delta)c(x,\eta)\pi(x) \\ &\geq 0\ . \end{aligned}$$

If moreover, condition (iii) is satisfied, the latter estimate can be improved by keeping only the terms with $\eta = \delta$ in the last sum. We thus obtain

$$\begin{aligned} \mathcal{B}(\rho,\psi) &\geq \frac{C}{4}\sum_{x,\delta} \big(2\nabla_\delta\psi(x)\big)^2\hat\rho(x,\delta x)c(x,\delta)\pi(x) \\ &= 2C\mathcal{A}(\rho,\psi)\ . \end{aligned}$$

□

Let us now consider some examples to which Proposition 5.4 can be applied.

Example 5.6 (The discrete circle). Consider the simple random walk on the discrete circle $C_n = \mathbb{Z}/n\mathbb{Z}$ of n sites given by the transition kernel $K(m, m-1) = K(m, m+1) = \frac{1}{2}$ for $m \in C_n$. We have the following mapping representation for K. Set $G = \{+, -\}$ where $+(m) = m+1$ and $-(m) = m-1$ and let $c(m,+) = c(m,-) = \frac{1}{2}$ for all m. Proposition 5.4 immediately yields that $\mathrm{Ric}(K) \geq 0$.

Example 5.7 (The discrete hypercube). Let $\mathcal{Q}^n = \{0,1\}^n$ be the hypercube endowed with the usual graph structure and let K_n be the kernel of the simple random walk on $\mathcal{Q}^n$. The natural mapping representation is given by $G = \{\delta_1, \ldots, \delta_n\}$, where $\delta_i : \mathcal{Q}^n \to \mathcal{Q}^n$ is the map that flips the i-th coordinate, and $c(x, \delta_i) = \frac{1}{n}$ for all $x \in \mathcal{Q}^n$. Here the criterion from Proposition 5.4 yields $\mathrm{Ric}(K_n) \geq \frac{2}{n}$. We shall see in Section 7 that this bound is optimal.

Alternatively, we can use the fact that $\mathcal{Q}^n$ is a product space and use the tensorisation property Theorem 6.2 below. This will allow to consider asymmetric random walks on the hypercube as well.

6. Basic Constructions

In this section we show how non-local Ricci curvature bounds transform under some basic operations on a Markov kernel. The main result is Theorem 6.2, which yields Ricci bounds for product chains. We start with a simple result that shows how Ricci bounds behave when adding laziness.

Let K be an irreducible and reversible Markov kernel on a finite set $\mathcal{X}$. For $\lambda \in (0,1)$ we consider the *lazy* Markov kernel defined by $K_\lambda := (1-\lambda)I + \lambda K$. Clearly, K_λ is irreducible and reversible with the same invariant measure π. With this notation, we have the following result:

Proposition 6.1 (Laziness)**.** *Let* $\lambda \in (0,1)$. *If* $\mathrm{Ric}(K) \geq \kappa$ *for some* $\kappa \in \mathbb{R}$, *then the lazy kernel* K_λ *satisfies*

$$\mathrm{Ric}(K_\lambda) \geq \lambda\kappa \ .$$

Proof. Writing $\mathcal{A}_\lambda$ and $\mathcal{B}_\lambda$ to denote the lazy versions of $\mathcal{A}$ and $\mathcal{B}$, a direct calculation shows that

$$\mathcal{A}_\lambda(\rho, \psi) = \lambda \mathcal{A}(\rho, \psi) \ , \qquad \mathcal{B}_\lambda(\rho, \psi) = \lambda^2 \mathcal{B}(\rho, \psi)$$

for all $\rho \in \mathscr{P}_*(\mathcal{X})$ and $\psi \in \mathbb{R}^{\mathcal{X}}$. As a consequence,

$$\mathcal{B}_\lambda(\rho, \psi) - \lambda\kappa\mathcal{A}_\lambda(\rho, \psi) = \lambda^2(\mathcal{B}(\rho, \psi) - \kappa\mathcal{A}(\rho, \psi)) \ .$$

The result thus follows from Theorem 4.5. □

We now give a tensorisation property of lower Ricci bounds with respect to products of Markov chains. For $i = 1, \ldots, n$, let $(\mathcal{X}_i, K_i)$ be an irreducible, reversible finite Markov chain with steady state π_i, and let α_i be a non-negative number satisfying $\sum_{i=1}^n \alpha_i = 1$. The product chain K_α on the product space $\mathcal{X} = \prod_i \mathcal{X}_i$ is defined for $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$

by

$$K_\alpha(\mathbf{x},\mathbf{y}) \;=\; \begin{cases} \sum_{i=1}^n \alpha_i K_i(x_i,x_i)\;, & \text{if } x_i = y_i \ \forall i\;,\\ \alpha_i K_i(x_i,y_i)\;, & \text{if } x_i \neq y_i \text{ and } x_j = y_j \ \forall j \neq i\;,\\ 0\;, & \text{otherwise}\;. \end{cases}$$

Note that the steady state of K_α is the product $\pi = \pi_1 \otimes \cdots \otimes \pi_n$ of the steady states of K_i.

Theorem 6.2 (Tensorisation)**.** *Assume that* $\mathrm{Ric}(K_i) \geq \kappa_i$ *for* $i = 1, \ldots, n$. *Then we have*

$$\mathrm{Ric}(K_\alpha) \;\geq\; \min_i \alpha_i \kappa_i\;.$$

Proof. In view of Theorem 4.5 we have to show that for any $\rho \in \mathscr{P}_*(\mathcal{X})$ and $\psi : \mathcal{X} \to \mathbb{R}$:

$$\mathcal{B}(\rho,\psi) \;\geq\; (\min_i \alpha_i\kappa_i)\mathcal{A}(\rho,\psi)\;.$$

We will use a mapping representation for the Markov kernel K_α as introduced in Section 5. Let (G_i, c_i) be mapping representations of K_i for $i = 1, \ldots, n$. To each $\delta \in G_i$ we associate a map $\bar{\delta} : \mathcal{X} \to \mathcal{X}$ by letting δ act on the i-th coordinate. Let us set $G = \bigcup_i \{\bar{\delta} : \delta \in G_i\}$ and define $c : \mathcal{X} \times G \to \mathbb{R}_+$ by

$$c(x,\bar{\delta}) \;:=\; \alpha_i c_i(x_i,\delta)\;, \quad \text{for } \delta \in G_i\;.$$

One easily checks that (G,c) is a mapping representation of K_α. Recalling the expressions (5.4),(5.5) which constitute $\mathcal{B}$ in mapping representation, we write

$$\mathcal{B}(\rho,\psi) \;=:\; \sum_{x\in\mathcal{X},\delta,\eta\in G} F(x,\delta,\eta)\;.$$

Taking into account the product structure of the chain we can write

$$\mathcal{B}(\rho,\psi) \;=\; \sum_{i,j=1}^n \mathcal{B}_{i,j} \quad \text{with} \quad \mathcal{B}_{i,j} \;=\; \sum_{x\in\mathcal{X}} \sum_{\delta\in G_i,\eta\in G_j} F(x,\bar{\delta},\bar{\eta})\;.$$

The proof will be finished if we prove the following two assertions:

(i) $\mathcal{B}_{i,j} \geq 0$ for all $i \neq j$,

(ii) $\sum_{i=1}^n \mathcal{B}_{i,i} \geq (\min_i \alpha_i\kappa_i)\mathcal{A}(\rho,\psi)$.

To show (i), first note that for $\delta \in G_i$ and $\eta \in G_j$ the maps $\bar{\delta}$ and $\bar{\eta}$ act on different coordinates if $i \neq j$. Thus we have $\bar{\delta} \circ \bar{\eta} = \bar{\eta} \circ \bar{\delta}$ and furthermore $c(\bar{\delta}x,\bar{\eta}) = c(x,\bar{\eta})$. Note that these are precisely the properties used in the proof Proposition 5.4, hence the assertion here follows from the same arguments.

Let us now show (ii). We set $\check{\mathcal{X}}_i = \prod_{j\neq i} \mathcal{X}_j$. For $\check{x}_i \in \check{\mathcal{X}}_i$ we let $\rho^{\check{x}_i}, \psi^{\check{x}_i} : \mathcal{X}_i \to \mathbb{R}$ denote the functions ρ and ψ where all variables except x_i are fixed to $\check{x}_i$. Note that $\rho^{\check{x}_i}$ does not necessarily belong to $\mathscr{P}(\mathcal{X}_i)$, but this will be irrelevant in the calculation below, and we shall use expressions such as

$\mathcal{A}(\rho^{\check{x}_i}, \psi^{\check{x}_i})$ by abuse of notation. We also set $\check{\pi}_i = \bigotimes_{j\neq i} \pi_j$. Once more using the product structure of the chain c, we see:

$$\begin{aligned}
&\mathcal{A}(\rho, \psi) \\
&= \frac{1}{2}\sum_{i=1}^{n} \sum_{x\in\mathcal{X},\delta\in G_i} \big(\nabla_{\bar{\delta}}\psi(x)\big)^2 \hat{\rho}(x, \bar{\delta}x)c(x, \bar{\delta})\pi(x) \\
&= \frac{1}{2}\sum_{i=1}^{n} \sum_{\check{x}_i\in\check{\mathcal{X}}_i} \sum_{x_i\in\mathcal{X}_i,\delta\in G_i} \big(\nabla_{\delta}\psi^{\check{x}_i}(x_i)\big)^2 \widehat{\rho^{\check{x}_i}}(x_i, \delta x_i)\alpha_i c_i(x_i, \delta)\pi_i(x_i)\check{\pi}_i(\check{x}_i) \\
&= \sum_{i=1}^{n} \alpha_i \sum_{\check{x}_i\in\check{\mathcal{X}}_i} \mathcal{A}_i(\rho^{\check{x}_i}, \psi^{\check{x}_i})\check{\pi}_i(\check{x}_i) \ ,
\end{aligned}$$

where $\mathcal{A}_i$ (resp. $\mathcal{B}_i$) denotes the function $\mathcal{A}$ (resp. $\mathcal{B}$) associated with the ith chain. Similarly, we obtain

$$\begin{aligned}
\mathcal{B}_{i,i} &= \alpha_i^2 \sum_{\check{x}_i\in\check{\mathcal{X}}_i} \mathcal{B}_i(\rho^{\check{x}_i}, \psi^{\check{x}_i})\check{\pi}_i(\check{x}_i) \\
&\geq \alpha_i^2 \kappa_i \sum_{\check{x}_i\in\check{\mathcal{X}}_i} \mathcal{A}_i(\rho^{\check{x}_i}, \psi^{\check{x}_i})\check{\pi}_i(\check{x}_i) \ ,
\end{aligned}$$

where the last inequality holds by assumption on the curvature bound for K_i. Summing over $i = 1, \ldots, n$ we obtain (ii). □

We shall now apply Theorem 6.2 to asymmetric random walks on the discrete hypercube. For $p, q \in (0, 1)$ let $K_{p,q}$ be the Markov kernel on the two point space $\{0, 1\}$ defined by $K(0, 1) = p, K(1, 0) = q$. The asymmetric random walk is the n-fold product chain on $\mathcal{Q}^n$ denoted by $K_{p,q,n}$ where $\alpha_i = \frac{1}{n}$. Note that the steady state of $K_{p,q,n}$ is the Bernoulli measure

$$\Big((1-\lambda)\delta_{\{0\}} + \lambda\delta_{\{1\}}\Big)^{\otimes n}$$

with parameter $\lambda = \frac{p}{p+q}$. We then have the following bound on the non-local Ricci curvature:

Proposition 6.3. *For $n \geq 1$ we have*

$$\mathrm{Ric}(K_{p,q,n}) \ \geq \ \frac{1}{n}\left(\frac{p+q}{2} + \sqrt{pq}\right) \ .$$

Proof. The two-point space $\mathcal{Q}^1 = \{0, 1\}$ has been analysed in detail in [28]. In particular, [28, Proposition 2.12] asserts that $\mathrm{Ric}(K_{p,q,1}) \geq \kappa_{p,q,n}$, where

$$\kappa_{p,q,n} \ = \ \frac{p+q}{2} + \inf_{-1<\beta<1}\left\{\frac{1}{1-\beta^2}\frac{q(1+\beta)-p(1-\beta)}{\log q(1+\beta) - \log p(1-\beta)}\right\} \ .$$

In order to estimate the right-hand side, we use the logarithmic-geometric mean inequality to obtain for $\beta \in (-1, 1)$,

$$\frac{1}{1-\beta^2}\frac{q(1+\beta)-p(1-\beta)}{\log q(1+\beta) - \log p(1-\beta)} \ \geq \ \sqrt{\frac{pq}{1-\beta^2}} \ \geq \ \sqrt{pq}$$

We thus infer that $\mathrm{Ric}(K_{p,q,1}) \geq \frac{p+q}{2} + \sqrt{pq}$. The general bound then follows immediately from Theorem 6.2. □

We shall see in Section 7 that this bound is sharp if $p = q$. If $p \neq q$, it should be possible to improve this bound by obtaining a sharper bound in the minimisation problem in the proof above.

As another application of the tensorisation result, we prove nonnegativity of the non-local Ricci curvature for the simple random walk on a discrete torus of arbitrary size in any dimension $d \geq 1$.

Let $\mathbf{c} := \{c_n\}_{n=1}^d$ be a sequence of natural numbers and consider the discrete torus

$$T_{\mathbf{c}} := C_{c_1} \times \ldots \times C_{c_d} \ .$$

The simple random walk $K_{\mathbf{c}}$ on $T_{\mathbf{c}}$ is the d-fold product of simple random walks on the circles of length $c_1, \ldots, c_d$.

Proposition 6.4 (d-dimensional torus)**.** *For any $d \geq 1$ and $\mathbf{c} := \{c_n\}_{n=1}^d \in \mathbb{N}^d$ we have*

$$\mathrm{Ric}(K_{\mathbf{c}}) \geq 0 \ .$$

Proof. This follows from Example 5.6 and Theorem 6.2. □

7. Functional Inequalities

The aim of this section is to prove discrete counterparts to the celebrated theorems by Bakry–Émery and Otto–Villani. Along the way we prove a discrete version of the HWI-inequality, which relates the L^2-Wasserstein distance to the entropy and the Fisher information. As announced in the introduction, we shall follow the approach from Otto–Villani, which relies on the fact that the heat flow is the gradient flow of the entropy. Therefore, the role of the L^2-Wasserstein distance will be taken over by the distance $\mathcal{W}$.

We fix a finite set $\mathcal{X}$ and an irreducible and reversible Markov kernel K with steady state π. Recall that the relative entropy of a density $\rho \in \mathscr{P}(\mathcal{X})$ is defined by

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

As before, we consider a discrete analogue of the Fisher information, given for $\rho \in \mathscr{P}_*(\mathcal{X})$ by

$$\mathcal{I}(\rho) \ = \ \frac{1}{2} \sum_{x,y \in \mathcal{X}} \big(\rho(x) - \rho(y)\big)\big(\log \rho(x) - \log \rho(y)\big) K(x,y) \pi(x) \ .$$

If $\rho(x) = 0$ for some $x \in \mathcal{X}$, we set $\mathcal{I}(\rho) = +\infty$. Note that this quantity can be rewritten in the form $\mathcal{I}(\rho) = \|\nabla \log \rho\|_\rho^2$ using the definition of the logarithmic mean. The relevance of $\mathcal{I}$ in this setting is due to the fact that it describes the entropy dissipation along the heat flow:

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathcal{H}(P_t \rho) \ = \ -\mathcal{I}(P_t \rho) \ . \tag{7.1}$$

The following proposition gives an upper bound for the speed of the heat flow measured in the metric $\mathcal{W}$.

Proposition 7.1. *Let $\rho, \sigma \in \mathscr{P}(\mathcal{X})$. For all $t > 0$ we have*

$$\frac{\mathrm{d}^+}{\mathrm{d}t}\mathcal{W}(P_t\rho, \sigma) \ \leq\ \sqrt{\mathcal{I}(P_t\rho)}\ . \tag{7.2}$$

In particular, the metric derivative of the heat flow with respect to $\mathcal{W}$ satisfies $|(P_t\rho)'| \leq \sqrt{\mathcal{I}(P_t\rho)}$. If ρ belongs to $\mathscr{P}_*(\mathcal{X})$, then (7.2) holds at $t = 0$ as well.

Proof. Let us set $\rho_t := P_t\rho$. Elementary Markov chain theory guarantees that $\rho_t \in \mathscr{P}_*(\mathcal{X})$ for all $t > 0$ and that the map $t \mapsto \rho_t$ is smooth. To prove (7.2) we use the triangle inequality and obtain

$$\begin{aligned}\frac{\mathrm{d}^+}{\mathrm{d}t}\mathcal{W}(\rho_t, \sigma) \ &=\ \limsup_{s\searrow 0}\frac{1}{s}\Big(\mathcal{W}(\rho_{t+s}, \sigma) - \mathcal{W}(\rho_t, \sigma)\Big)\\ &\leq\ \limsup_{s\searrow 0}\frac{1}{s}\mathcal{W}(\rho_t, \rho_{t+s})\ .\end{aligned}$$

Note that the couple $(\rho_r, -\log\rho_r)_{r\in[0,1]}$ solves the continuity equation (1.2). From the definition of $\mathcal{W}$ we thus obtain the estimate

$$\begin{aligned}\limsup_{s\searrow 0}\frac{1}{s}\mathcal{W}(\rho_t, \rho_{t+s}) \ &\leq\ \limsup_{s\searrow 0}\frac{1}{s}\int_t^{t+s} \|\ \nabla\log\rho_r\ \|_{\rho_r}\ \mathrm{d}r\\ &=\ \limsup_{s\searrow 0}\frac{1}{s}\int_t^{t+s}\sqrt{\mathcal{I}(\rho_r)}\,\mathrm{d}r\\ &=\ \sqrt{\mathcal{I}(\rho_t)}\ .\end{aligned}$$

The last equality holds since $r \mapsto \sqrt{\mathcal{I}(\rho_r)}$ is a continuous function. □

Let us now recall from Section 1 the functional inequalities that will be studied. Recall that $\mathbf{1} \in \mathscr{P}(\mathcal{X})$ denotes the density of the stationary distribution, which is everywhere equal to 1.

Definition 7.2. *The Markov kernel K satisfies*

(1) a modified logarithmic Sobolev *inequality with constant $\lambda > 0$ if for all $\rho \in \mathscr{P}(\mathcal{X})$,*

$$\mathcal{H}(\rho) \ \leq\ \frac{1}{2\lambda}\mathcal{I}(\rho)\ . \tag{MLSI(λ)}$$

(2) an H$\mathcal{W}$I *inequality with constant $\kappa \in \mathbb{R}$ if for all $\rho \in \mathscr{P}(\mathcal{X})$,*

$$\mathcal{H}(\rho) \ \leq\ \mathcal{W}(\rho, \mathbf{1})\sqrt{\mathcal{I}(\rho)} - \frac{\kappa}{2}\mathcal{W}(\rho, \mathbf{1})^2\ . \tag{H$\mathcal{W}$I(κ)}$$

(3) a modified Talagrand *inequality with constant $\lambda > 0$ if for all $\rho \in \mathscr{P}(\mathcal{X})$,*

$$\mathcal{W}(\rho, \mathbf{1}) \ \leq\ \sqrt{\frac{2}{\lambda}\mathcal{H}(\rho)}\ . \tag{T$_{\mathcal{W}}$(λ)}$$

(4) a Poincaré *inequality with constant $\lambda > 0$ if for all $\varphi \in \mathbb{R}^X$ with $\sum_x \varphi(x)\pi(x) = 0$,*

$$\|\varphi\|_\pi^2 \ \leq\ \frac{1}{\lambda}\|\nabla\varphi\|_\pi^2\ . \tag{P(λ)}$$

The following result is a discrete analogue of a result by Otto and Villani [37].

Theorem 7.3. *Assume that* $\mathrm{Ric}(K) \geq \kappa$ *for some* $\kappa \in \mathbb{R}$*. Then* K *satisfies* $\mathrm{H}\mathcal{W}\mathrm{I}(\kappa)$*.*

Proof. Fix $\rho \in \mathscr{P}(\mathcal{X})$. Without restriction we can assume that $\rho > 0$ since otherwise $\mathcal{I}(\rho) = +\infty$ and there is nothing to prove. Let $\rho_t = P_t\rho$ where $P_t = e^{t(K-I)}$ is the heat semigroup. From Theorem 4.5 and the lower bound on the Ricci curvature we know that the curve (ρ_t) satisfies $\mathrm{EVI}(\kappa)$, i.e., equation (4.4). Choosing, in particular, $\nu = \mathbf{1}$ and $t = 0$ in the EVI we obtain the inequality

$$\mathcal{H}(\rho) \ \leq \ -\frac{1}{2}\left.\frac{\mathrm{d}^+}{\mathrm{d}t}\right|_{t=0} \mathcal{W}(\rho_t, \mathbf{1})^2 - \frac{\kappa}{2}\mathcal{W}(\rho, \mathbf{1})^2 \ .$$

To finish the proof we show that

$$-\frac{1}{2}\left.\frac{\mathrm{d}^+}{\mathrm{d}t}\right|_{t=0} \mathcal{W}(\rho_t, \mathbf{1})^2 \ \leq \ \mathcal{W}(\rho, \mathbf{1})\sqrt{\mathcal{I}(\rho)} \ .$$

Indeed, using the triangle inequality we estimate

$$\begin{aligned} -\frac{1}{2}\left.\frac{\mathrm{d}^+}{\mathrm{d}t}\right|_{t=0} \mathcal{W}(\rho_t, \mathbf{1})^2 \ &= \ \liminf_{s\searrow 0} \frac{1}{2s}\big(\mathcal{W}(\rho, \mathbf{1})^2 - \mathcal{W}(\rho_s, \mathbf{1})^2\big) \\ &\leq \ \limsup_{s\searrow 0} \frac{1}{2s}\big(\mathcal{W}(\rho, \rho_s)^2 + 2\mathcal{W}(\rho, \rho_s)\cdot \mathcal{W}(\rho, \mathbf{1})\big) \ , \end{aligned}$$

Using the estimate (7.2) from Proposition 7.1 with $\sigma = \rho$ and $t = 0$ we see that the second term on the right-hand side is bounded by $\mathcal{W}(\rho, \mathbf{1})\sqrt{\mathcal{I}(\rho)}$ while the first term vanishes. □

The following result is now a simple consequence.

Theorem 7.4 (Discrete Bakry–Émery Theorem)**.** *Assume that* $\mathrm{Ric}(K) \geq \lambda$ *for some* $\lambda > 0$*. Then* K *satisfies* $\mathrm{MLSI}(\lambda)$*.*

Proof. By Theorem 7.3 K satisfies $\mathrm{H}\mathcal{W}\mathrm{I}(\lambda)$. From this we derive $\mathrm{MLSI}(\lambda)$ by an application of Young's inequality :

$$xy \ \leq \ cx^2 + \frac{1}{4c}y^2 \qquad \forall x, y \in \mathbb{R}\ , \ c > 0\ ,$$

in which we set $x = \mathcal{W}(\rho, 1)$, $y = \sqrt{\mathcal{I}(\rho)}$ and $c = \frac{\lambda}{2}$. □

Theorem 7.5 (Discrete Otto–Villani Theorem)**.** *Assume that* K *satisfies* $\mathrm{MLSI}(\lambda)$ *for some* $\lambda > 0$*. Then* K *also satisfies* $\mathrm{T}_{\mathcal{W}}(\lambda)$*.*

Proof. It is sufficient to prove that $\mathrm{T}_{\mathcal{W}}(\lambda)$ holds for any $\rho \in \mathscr{P}_*(\mathcal{X})$. The inequality for general ρ can then be obtained by an easy approximation argument taking into account the continuity of $\mathcal{W}$ with respect to the Euclidean metric.

So, fix $\rho \in \mathscr{P}_*(\mathcal{X})$ and set $\rho_t = P_t\rho$. First note that as $t \to \infty$, we have

$$\mathcal{H}(\rho_t) \to 0 \quad \text{ and } \quad \mathcal{W}(\rho, \rho_t) \to \mathcal{W}(\rho, \mathbf{1}) \ . \tag{7.3}$$

Indeed, by elementary Markov chain theory, we know that as $t \to \infty$, one has $\rho_t \to \mathbf{1}$ in, say, the Euclidean distance. The claim follows immediately from the continuity of $\mathcal{H}$ and $\mathcal{W}$ with respect to the Euclidean distance, the latter being a consequence of, for example, Proposition 2.14.

We now define the function $F : \mathbb{R}_+ \to \mathbb{R}_+$ by

$$F(t) \;:=\; \mathcal{W}(\rho, \rho_t) + \sqrt{\frac{2}{\lambda}\mathcal{H}(\rho_t)}\;.$$

Obviously we have $F(0) = \sqrt{\frac{2}{\lambda}\mathcal{H}(\rho)}$ and by (7.3) we have that $F(t) \to \mathcal{W}(\rho, \mathbf{1})$ as $t \to \infty$. Hence it is sufficient to show that F is non-increasing. To this end we show that its upper right derivative is non-positive. If $\rho_t \neq \mathbf{1}$ we deduce from Proposition 7.1 that

$$\frac{\mathrm{d}^+}{\mathrm{d}t}F(t) \;\leq\; \sqrt{\mathcal{I}(\rho_t)} - \frac{\mathcal{I}(\rho_t)}{\sqrt{2\lambda\mathcal{H}(\rho_t)}} \;\leq\; 0\;,$$

where we used MLSI(λ) in the last inequality. If $\rho_t = \mathbf{1}$, then the relation also holds true, since this implies that $\rho_r = \mathbf{1}$ for all $r \geq t$. □

In a classical continuous setting it is well known that a logarithmic Sobolev inequality implies a Poincaré inequality by linearisation. Let us make this explicit in the present discrete context. Fix $\varphi \in \mathbb{R}^{\mathcal{X}}$ satisfying $\sum_x \varphi(x)\pi(x) = 0$ and for sufficiently small $\varepsilon > 0$ set $\rho^\varepsilon = \mathbf{1} + \varepsilon\varphi \in \mathscr{P}_*(\mathcal{X})$. One easily checks that as $\varepsilon \to 0$ we have:

$$\frac{1}{\varepsilon^2}\mathcal{H}(\rho^\varepsilon) \longrightarrow \frac{1}{2}\|\varphi\|_\pi^2\;, \qquad \frac{1}{\varepsilon^2}\mathcal{I}(\rho^\varepsilon) \longrightarrow \|\nabla\varphi\|_\pi^2\;.$$

Thus assuming MLSI(λ) holds and applying it to ρ^ε we get the Poincaré inequality P(λ). In [37] it has been shown that the Poincaré inequality can also be obtained from Talagrand's inequality by linearisation. The same is true for the modified Talagrand inequality involving the distance $\mathcal{W}$.

Proposition 7.6. *Assume that K satisfies $T_{\mathcal{W}}(\lambda)$ for some $\lambda > 0$. Then K also satisfies $P(\lambda)$. In particular,* $\mathrm{Ric}(K) \geq \lambda$ *implies* P(λ).

Proof. Assume that $\mathrm{T}_{\mathcal{W}}(\lambda)$ holds and let us show P(λ). The second assertion of the proposition then follows from Theorem 7.4 and Theorem 7.5. So fix $\varphi \in \mathbb{R}^{\mathcal{X}}$ satisfying $\sum_x \varphi(x)\pi(x) = 0$ and for sufficiently small $\varepsilon > 0$ set $\rho^\varepsilon = \mathbf{1} + \varepsilon\varphi \in \mathscr{P}_*(\mathcal{X})$. Let $(\rho^\varepsilon_\cdot, V^\varepsilon_\cdot) \in \mathcal{CE}_1'(\rho^\varepsilon, \mathbf{1})$ be an action minimizing curve. Now we write, using the continuity equation,

$$\begin{aligned}\sum_x \varphi(x)^2\pi(x) \;&=\; \frac{1}{\varepsilon}\left[\sum_x \varphi(x)\big(\rho^\varepsilon(x) - 1\big)\pi(x)\right]\\ &=\; \frac{1}{2\varepsilon}\int_0^1 \sum_{x,y} \nabla\varphi(x,y)V_t^\varepsilon(x,y)K(x,y)\pi(x)\,\mathrm{d}t\;.\end{aligned}$$

Using Hölder's inequality we can estimate

$$\begin{aligned}\sum_x \varphi(x)^2\pi(x) \;&\leq\; \frac{1}{\varepsilon}\left(\int_0^1 \|\nabla\varphi\|_{\rho_t^\varepsilon}^2\,\mathrm{d}t\right)^{\frac{1}{2}}\left(\int_0^1 \mathcal{A}'(\rho_t^\varepsilon, V_t^\varepsilon)\,\mathrm{d}t\right)^{\frac{1}{2}}\\ &=\; \left(\frac{1}{2}\sum_{x,y}\big(\nabla\varphi(x,y)\big)^2 g^\varepsilon(x,y)K(x,y)\pi(x)\right)^{\frac{1}{2}}\frac{1}{\varepsilon}\mathcal{W}(\rho^\varepsilon, \mathbf{1})\;,\end{aligned}$$

where $g^\varepsilon \in \mathbb{R}^{\mathcal{X}\times\mathcal{X}}$ is defined by $g^\varepsilon(x,y) = \int_0^1 \hat\rho_t^\varepsilon(x,y)\,\mathrm{d}t$. Using $\mathrm{T}_{\mathcal{W}}(\lambda)$ we arrive at

$$\|\varphi\|_\pi^2 \ \le\ \|(\nabla\varphi)\sqrt{g^\varepsilon}\|_\pi \frac{1}{\varepsilon}\sqrt{\frac{2}{\lambda}\mathcal{H}(\rho^\varepsilon)}\ .$$

The proof will be finished if we show that as ε goes to 0

$$\frac{1}{\varepsilon}\sqrt{\frac{2}{\lambda}\mathcal{H}(\rho^\varepsilon)} \ \longrightarrow\ \sqrt{\frac{1}{\lambda}}\|\varphi\|_\pi\ , \qquad \|(\nabla\varphi)\sqrt{g^\varepsilon}\|_\pi \ \longrightarrow\ \|\nabla\varphi\|_\pi\ .$$

As before, the first statement is easily checked. For the second statement it is sufficient to show that $\rho_t^\varepsilon \to \mathbf{1}$ uniformly in t as $\varepsilon\to 0$, as this implies that $g^\varepsilon \to 1$. Since $\mathcal{W}(\rho^\varepsilon,\mathbf{1})\to 0$ as $\varepsilon\to 0$, this follows immediately from the estimate

$$\mathcal{W}(\rho^\varepsilon,\mathbf{1}) \ \ge\ \sup_t \mathcal{W}(\rho_t^\varepsilon,\mathbf{1}) \ \ge\ \sup_t \sum_x \pi(x)|\rho_t^\varepsilon(x)-1|\ ,$$

where we used that $(\rho_t^\varepsilon)_{t\in[0,1]}$ is a geodesic and the fact that $\mathcal{W}$ is an upper bound for the total variation distance (see Proposition 2.12). □

In the following result we use the probabilistic notation

$$\mathbb{E}_\pi[\varphi] = \sum_{x\in\mathcal{X}} \varphi(x)\pi(x)$$

for functions $\varphi:\mathcal{X}\to\mathbb{R}$.

Proposition 7.7. *Assume that K satisfies $\mathrm{T}_{\mathcal{W}}(\lambda)$ for some $\lambda>0$. Then the $T_1(2\lambda)$ inequality holds with respect to the graph distance:*

$$W_{1,g}(\rho,\mathbf{1}) \ \le\ \sqrt{\frac{1}{\lambda}\mathcal{H}(\rho)}\ . \tag{7.4}$$

Furthermore, the sub-Gaussian inequality

$$\mathbb{E}_\pi\Big[e^{t(\varphi-\mathbb{E}_\pi[\varphi])}\Big] \le \exp\Big(\frac{t^2}{4\lambda}\Big) \tag{7.5}$$

holds for all $t>0$ and every function $\varphi:\mathcal{X}\to\mathbb{R}$ that is 1-Lipschitz with respect to the graph distance on $\mathcal{X}$.

Proof. The T_1-inequality (7.4) follows immediately from Proposition 2.12. The inequalities (7.4) and (7.5) are equivalent, as has been shown in [8]. □

Arguing again exactly as in [37], we infer that a modified Talagrand inequality implies a modified log-Sobolev inequality (with some loss in the constant), provided that the non-local Ricci curvature is not too bad.

Proposition 7.8. *Suppose that K satisfies $\mathrm{T}_{\mathcal{W}}(\lambda)$ for some $\lambda>0$ and that $\mathrm{Ric}(K)\ge\kappa$ for some $\kappa>-\lambda$. Then K satisfies $\mathrm{MLSI}(\widetilde\lambda)$, where*

$$\widetilde\lambda = \max\left\{\frac{\lambda}{4}\Big(1+\frac{\kappa}{\lambda}\Big)^2\ ,\ \kappa\right\}.$$

Proof. This is an immediate consequence of the $\mathrm{H}\mathcal{W}\mathrm{I}(\kappa)$-inequality and an elementary computation (see [37, Corollary 3.1]). □

As an application of the results proved in this section, we will show how non-local Ricci curvature bounds can be used to recover functional inequalities with sharp constants in an important example.

Example 7.9 (Discrete hypercube). In Example 5.7 and Proposition 6.3 we proved that the Markov kernel K_n associated with the simple random walk on the discrete hypercube $\mathcal{Q}^n = \{0,1\}^n$ has non-local Ricci curvature bounded from below by $\frac{2}{n}$. Applying Theorem 7.4 and Proposition 7.7 in this setting, we obtain the following result. We shall write $y \sim x$ if $K(x,y) > 0$.

Corollary 7.10. *The simple random walk on $\mathcal{Q}^n$ has the following properties:*

(1) the modified log-Sobolev inequality $\mathrm{MLSI}(\frac{2}{n})$ *holds, that is, for all $\rho \in \mathscr{P}_*(\mathcal{Q}^n)$ we have*

$$\sum_{x\in\mathcal{Q}^n} \rho(x)\log\rho(x) \;\le\; \frac{1}{8}\sum_{x\in\mathcal{Q}^n, y\sim x} (\rho(x)-\rho(y))\,(\log\rho(x)-\log\rho(y)) \;.$$

(2) the Poincaré inequality $\mathrm{P}(\frac{2}{n})$ *holds, that is, for all $\varphi : \mathcal{Q}^n \to \mathbb{R}$ we have*

$$\sum_{x\in\mathcal{Q}^n} \varphi(x)^2 \;\le\; \frac{1}{4}\sum_{x\in\mathcal{Q}^n, y\sim x} (\varphi(x)-\varphi(y))^2 \;.$$

(3) The sub-Gaussian inequality (7.5) *holds with $\lambda = \frac{2}{n}$.*

In all cases the constants are optimal (see [11, Example 3.7] and [9, Proposition 2.3] respectively). Moreover, the optimality in (3) implies that the constant $\lambda = \frac{2}{n}$ in the modified Talagrand inequality for the discrete cube is sharp, as well.

We finish the paper by remarking that modified logarithmic Sobolev inequalities for appropriately rescaled product chains on the discrete hypercube $\{-1,1\}^n$ can be used to prove a similar inequality for Poisson measures by passing to the limit $n \to \infty$ (see [25, Section 5.4] for an argument along these lines involving a slightly different modified log Sobolev inequality). All of the functional inequalities in Theorem 1.5 are compatible with this limit. However, the sub-Gaussian estimate will (of course) not hold for the limiting Poisson law. This does not contradict the results in this section, since the sub-Gaussian estimates here are obtained using the lower bound for $\mathcal{W}$ in terms of W_1, which relies on the normalisation assumption $\sum_y K(x,y) = 1$, which does not hold in the Poissonian limit.

References

[1] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[2] L. Ambrosio, N. Gigli, and G. Savaré. Calculus and heat flow in metric measure spaces and applications to spaces with Ricci bounds from below. *Preprint at arXiv:1106.2090*, 2011.

[3] L. Ambrosio, N. Gigli, and G. Savaré. Metric measure spaces with Riemannian Ricci curvature bounded from below. *Preprint at arXiv:1109.0222*, 2011.

[4] C. Ané and M. Ledoux. On logarithmic Sobolev inequalities for continuous time random walks on graphs. *Probab. Theory Related Fields*, 116(4):573–602, 2000.

[5] 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.

[6] F. Bauer, J. Jost, and S. Liu. Ollivier-Ricci curvature and the spectrum of the normalized graph Laplace operator. *Preprint at arXiv:1105.3803*, 2011.

[7] J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numer. Math.*, 84(3):375–393, 2000.

[8] S. G. Bobkov and F. Götze. Discrete isoperimetric and Poincaré-type inequalities. *Probab. Theory Related Fields*, 114(2):245–277, 1999.

[9] S. G. Bobkov, C. Houdré, and P. Tetali. The subgaussian constant and concentration inequalities. *Israel J. Math.*, 156:255–283, 2006.

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

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

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

[13] G. Buttazzo. *Semicontinuity, relaxation and integral representation in the calculus of variations*. Pitman Research Notes in Mathematics Series. Longman Scientific and Technical, Harlow, 1989.

[14] 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.

[15] 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. Rational Mech. Anal., Online First.*

[16] S. Daneri and G. Savaré. Eulerian calculus for the displacement convexity in the Wasserstein distance. *SIAM J. Math. Anal.*, 40(3):1104–1122, 2008.

[17] J. Dolbeault, B. Nazaret, and G. Savaré. A new class of transport distances between measures. *Calc. Var. Partial Differential Equations*, 34(2):193–231, 2009.

[18] M. Erbar. The heat equation on manifolds as a gradient flow in the Wasserstein space. *Ann. Inst. Henri Poincaré Probab. Stat.*, 46(1):1–23, 2010.

[19] S. Fang, J. Shao, and K.-Th. Sturm. Wasserstein space over the Wiener space. *Probab. Theory Related Fields*, 146(3-4):535–565, 2010.

[20] N. Gigli, K. Kuwada, and S.-I. Ohta. Heat flow on Alexandrov spaces. *Preprint at arXiv:1008.1319*, 2010.

[21] N. Gozlan and C. Léonard. Transport inequalities. A survey. *Markov Process. Related Fields*, 16(4):635–736, 2010.

[22] B. Hua, J. Jost, and S. Liu. Geometric analysis aspects of infinite semiplanar graphs with nonnegative curvature. *Preprint at arXiv:1107.2826*, 2011.

[23] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.

[24] J. Jost and S. Liu. Ollivier's Ricci curvature, local clustering and curvature dimension inequalities on graphs. *Preprint at arXiv:1103.4037*, 2011.

[25] M. Ledoux. *The concentration of measure phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2001.

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

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

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

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

[30] A. Mielke. Geodesic convexity of the relative entropy in reversible Markov chains. *Preprint*, 2011.

[31] A. Mielke. A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. *Nonlinearity*, 24(4):1329–1346, 2011.

[32] S.-I. Ohta and K.-Th. Sturm. Heat flow on Finsler manifolds. *Comm. Pure Appl. Math.*, 62(10):1386–1433, 2009.

[33] Y. Ollivier. Ricci curvature of metric spaces. *C. R. Math. Acad. Sci. Paris*, 345(11):643–646, 2007.

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

[35] Y. Ollivier and C. Villani. A curved Brunn-Minkowski inequality on the discrete hypercube. *Preprint at arXiv:1011.4779*, 2010.

[36] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

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

[38] F. Otto and M. Westdickenberg. Eulerian calculus for the contraction in the Wasserstein distance. *SIAM J. Math. Anal.*, 37(4):1227–1255 (electronic), 2005.

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

[40] M. Sammer and P. Tetali. Concentration on the discrete torus using transportation. *Combin. Probab. Comput.*, 18(5):835–860, 2009.

[41] K.-Th. Sturm. On the geometry of metric measure spaces. I and II. *Acta Math.*, 196(1):65–177, 2006.

[42] C. Villani. *Topics in optimal transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2003.

[43] C. Villani. *Optimal transport, Old and new*, volume 338 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, 2009.

University of Bonn, Institute for Applied Mathematics, Endenicher Allee 60, 53115 Bonn, Germany

E-mail address: erbar@iam.uni-bonn.de

E-mail address: maas@iam.uni-bonn.de