ENTROPIC RICCI CURVATURE FOR DISCRETE SPACES

JAN MAAS

Abstract. We give a short overview on a recently developed notion of Ricci curvature for discrete spaces. This notion relies on geodesic convexity properties of the relative entropy along geodesics in the space of probability densities, for a metric which is similar to (but different from) the 2-Wasserstein metric. The theory can be considered as a discrete counterpart to the theory of Ricci curvature for geodesic measure spaces developed by Lott–Sturm–Villani.

1. Ricci curvature lower bounds for geodesic measure spaces

In the last decade, ideas from the theory of optimal transport have led to significant progress in the analysis and geometry on non-smooth spaces. The starting point for these developments were the independent works by Lott and Villani [21] and Sturm [30], who introduced a notion of Ricci curvature for metric measure spaces based on a beautiful connection between optimal transport and entropy that originates in McCann's pioneering work [23].

Let $(\mathcal{X}, d)$ be a complete and separable metric space and let $\mathscr{P}(\mathcal{X})$ be the space of Borel probability measures on $\mathcal{X}$. For $1 \leq p < \infty$ and $\mu, \nu \in \mathscr{P}(\mathcal{X})$ we consider the *Monge-Kantorovich metric* W_p (often called Wasserstein metric), defined by

$$W_p(\mu, \nu) := \inf_{\Gamma \in \Pi(\mu,\nu)} \left(\int_{\mathcal{X} \times \mathcal{X}} d(x,y)^p \, \mathrm{d}\Gamma(x,y) \right)^{1/p} \qquad \text{for } \mu, \nu \in \mathscr{P}(\mathcal{X}) ,$$

where $\Pi(\mu, \nu)$ denotes the set of all couplings of μ and ν, i.e., all probability measures $\Gamma \in \mathscr{P}(\mathcal{X} \times \mathcal{X})$ satisfying

$$\Gamma(A \times \mathcal{X}) = \mu(A) \quad \text{and} \quad \Gamma(\mathcal{X} \times A) = \nu(A)$$

for all Borel sets $A \subseteq \mathcal{X}$. Loosely speaking, $W_p(\mu_0, \mu_1)^p$ is the minimal transportation cost required to transport an amount of mass from its initial configuration μ to a prescribed final configuration ν, at a cost of $d(x, y)^p$ per unit. It can be shown that W_p defines a metric on $\mathscr{P}_p(\mathcal{X})$, the space of probability measures with finite p-th moment. Moreover, if $(\mathcal{X}, d)$ is a geodesic space (i.e., every pair of points $x_0, x_1 \in \mathcal{X}$ can be joined by a curve $\gamma : [0, 1] \to \mathcal{X}$ such that $d(\gamma(s), \gamma(t)) = |s - t| d(x_0, x_1)$ for all $s, t \in [0, 1]$), then the 2-Wasserstein space $(\mathscr{P}_2(\mathcal{X}), W_2)$ is a geodesic space as well.

The final version of this manuscript will appear in *Modern Approaches to Discrete Curvature* (eds. L. Najman and P. Romon), Springer Lecture Notes in Mathematics.

Given a reference measure $\nu \in \mathscr{P}(\mathcal{X})$, the relative entropy with respect to ν is defined by

$$\mathrm{Ent}_\nu(\mu) := \int_{\mathcal{X}} \rho(x) \log \rho(x) \, \mathrm{d}\nu(x)$$

whenever $\mu \in \mathscr{P}(\mathcal{X})$ is absolutely continuous with density $\rho = \frac{\mathrm{d}\mu}{\mathrm{d}\nu}$, provided that the integral is well-defined. If ν is a probability measure, this quantity takes values in $[0, +\infty]$.

The following result, proved in [28, 9, 29], characterises Ricci curvature lower bounds on Riemannian manifolds in terms of convexity properties of the relative entropy (with respect to the volume measure) and optimal transport.

Theorem 1 (Characterisation of Ricci lower bounds on Riemannian manifolds)**.** *Let $\kappa \in \mathbb{R}$. For a complete Riemannian manifold $\mathcal{M}$, the following assertions are equivalent:*

(1) $\mathrm{Ric} \geq \kappa$ *on $\mathcal{M}$.*
(2) *Each pair of probability measures $\mu_0, \mu_1 \in \mathscr{P}_2(\mathcal{M})$ can be connected by a constant speed W_2-geodesic $(\mu_t)_{t\in[0,1]}$ along which the entropy satisfies the κ-convexity inequality*

$$\mathrm{Ent}_{\mathrm{vol}}(\mu_t) \leq (1-t)\,\mathrm{Ent}_{\mathrm{vol}}(\mu_0) + t\,\mathrm{Ent}_{\mathrm{vol}}(\mu_1) - \frac{\kappa}{2} t(1-t) W_2(\mu_0,\mu_1)^2 \ .$$

While the definition of the Ricci tensor requires a Riemannian structure on the underlying space, the second condition makes sense in much greater generality: the sole requirements are a metric (to define the Wasserstein metric) and a measure (to define the relative entropy). Therefore the following definition makes sense:

Definition 2 (Lott–Sturm–Villani)**.** *Let $\kappa \in \mathbb{R}$. A metric measure space $(\mathcal{X}, d, \nu)$ is said to have "Ricci curvature bounded from below by κ" if every pair of probability measures $\mu_0, \mu_1 \in \mathscr{P}_2(\mathcal{X})$ can be connected by a constant speed W_2-geodesic $(\mu_t)_{t\in[0,1]}$ along which the relative entropy Ent_ν satisfies the κ-convexity inequality*

$$\mathrm{Ent}_\nu(\mu_t) \leq (1-t)\,\mathrm{Ent}_\nu(\mu_0) + t\,\mathrm{Ent}_\nu(\mu_1) - \frac{\kappa}{2} t(1-t) W_2(\mu_0,\mu_1)^2 \ .$$

This definition has become the starting point for many interesting developments in the analysis and geometry on metric measure spaces. A large number of geometric and functional inequalities with sharp constants can be derived from this definition. A crucial feature of the theory is its robustness, i.e., the *stability* of lower Ricci curvature bounds under convergence of metric measure space in the sense of measured Gromov–Hausdorff convergence. A stronger version of Definition 2, a curvature-dimension criterion, has been introduced in [21, 30] as well. We refer to [31] for an overview of this theory. More recently, important refinements of the curvature-dimension criterion have been introduced in [2, 13].

1.1. **Discrete spaces.** If $(\mathcal{X}, d)$ is a geodesic space, the Lott–Sturm–Villani criterion is non-trivial, since we have already mentioned that the 2-Wasserstein space $(\mathscr{P}_2(\mathcal{X}), W_2)$ is a geodesic space as well. If $(\mathcal{X}, d)$ is *discrete*, the situation turns out to be completely different. Indeed, suppose that $(\mu_t)_{t\in[0,1]}$ is W_2-Lipschitz, i.e., $W_2(\mu_s, \mu_t) \leq L|s-t|$ for some $L < \infty$. Fix a point $x \in \mathcal{X}$ and set $m(t) := \mu_t(\{x\})$. Since $\mathcal{X}$ is discrete, there exists $\delta > 0$ such that all other points are at least at distance δ from x. As a consequence, $W_2(\mu_s,\mu_t) \geq \delta\sqrt{|m(t)-m(s)|}$. Combining the estimates above, we infer that $t \mapsto m(t)$ is Hölder continuous with exponent 2, hence m is constant. Since x was arbitrary, we conclude that every W_2-Lipschitz curve is constant. In particular, there are no W_2-geodesics, so that the Lott–Sturm–Villani criterion is (trivially) not satisfied for any $\kappa \in \mathbb{R}$.

While the curvature concept based on geodesic convexity of the entropy has been extremely powerful in the continuous (geodesic) setting, these observations show that the 2-Wasserstein metric is not the appropriate object for an analogous discrete theory.

To motivate the definition of a suitable discrete counterpart of W_2, we shall describe a seminal result in the continuous setting in which the Wasserstein metric plays a central role.

2. The heat flow as gradient flow of the entropy

It is a classical fact that the heat equation $\partial_t u = \Delta u$ can be regarded as the gradient flow equation in $L^2(\mathbb{R}^n)$ for the Dirichlet energy $\mathcal{E} : L^2(\mathbb{R}^n) \to \mathbb{R} \cup \{+\infty\}$ given by

$$\mathcal{E}(u) = \begin{cases} \frac{1}{2}\int_{\mathbb{R}^n} |\nabla u(x)|^2 \,\mathrm{d}x\ , & u \in H^1(\mathbb{R}^n)\ , \\ +\infty\ , & \text{otherwise.} \end{cases}$$

A quite different gradient flow structure for the heat equation, physically very appealing, was discovered by Jordan, Kinderlehrer, and Otto [19] at the end of the 1990s.

Theorem 3 (Heat flow is gradient flow of the entropy)**.** *The heat equation $\partial_t \rho = \Delta\rho$ is the gradient flow equation for the Boltzmann-Shannon entropy in the 2-Wasserstein space $(\mathscr{P}_2(\mathbb{R}^n), W_2)$.*

Note that the meaning of this statement is not clear a priori . While gradient flows are traditionally considered in smooth (e.g. Riemannian or Hilbertian) settings, the statement of this result involves the notion of a gradient flow in a metric space, in which the notion of a gradient is not defined.

In the original paper [19], the result was made rigorous using the discrete minimising movement scheme given by

$$\mu^{(0)} := \mu_0\ , \qquad \mu^{(k+1)} := \mathrm{argmin}_{\sigma\in\mathscr{P}(\mathbb{R}^n)} \left(\mathrm{Ent}(\sigma) + \frac{1}{2h} W_2(\mu^{(k)},\sigma)^2 \right)$$

The authors showed that the piecewise constant functions (μ_h) defined by

$$\mu_h(t) := \mu^{(k)}\ , \qquad t \in [kh,(k+1)h)\ ,$$

converge, as $h \downarrow 0$, to the solution of the heat equation $\partial_t\mu = \Delta\mu$ with initial condition $\mu|_{t=0} = \mu_0$. It is also possible to interpret Theorem 3 using the theory of gradient flows in metric spaces, which has been systematically developed in [1].

Another interpretation of this result can be given in terms of a formal infinite-dimensional Riemannian structure on the space of probability measures [27]. This structure is closely related to the so-called Benamou-Brenier formula, which provides a dynamical characterisation of the 2-Wasserstein metric inspired by fluid mechanics.

2.1. **The Benamou-Brenier formula.** Fix two probability measures $\mu_0, \mu_1 \in \mathscr{P}_2(\mathbb{R}^n)$. Instead of considering couplings as in the Monge–Kantorovich problem, we consider continuous-time interpolations $(\mu_t)_{t\in[0,1]}$ between μ_0 and μ_1. Under mild regularity conditions, such interpolations satisfy the continuity equation (in the sense of distributions)

$$\partial_t \mu_t + \nabla\cdot(\mu_t v_t) = 0 \tag{1}$$

for a suitable time-dependent velocity vector field $v : [0,1] \times \mathbb{R}^n \to \mathbb{R}^n$. The remarkable Benamou–Brenier formula [4] asserts that the squared Wasserstein distance is obtained by minimising the total kinetic energy among all solutions to the continuity equation. More precisely,

$$W_2(\mu_0, \mu_1)^2 = \inf_{(\mu_t, v_t)_t} \left\{ \int_0^1 \int_{\mathbb{R}^n} |v_t(x)|^2 \, \mathrm{d}\mu_t(x) \, \mathrm{d}t \right\} , \tag{2}$$

where the infimum runs over all sufficiently regular solutions $(\mu_t, v_t)_{t\in[0,1]}$ to (1) with boundary conditions $\mu|_{t=0} = \mu_0$ and $\mu|_{t=1} = \mu_1$.

Given a sufficiently regular curve of probability measures $(\mu_t)_{t\in(-\epsilon,\epsilon)}$, there are many different velocity vector fields v that satisfy the continuity equation (1) at $t = 0$. However, it turns out that only one of those vector fields is a *generalised gradient*, in the sense that it belongs to the closure of $\{\nabla\psi : \psi \in C_c^\infty(\mathbb{R}^n)\}$ in $L^2(\mathbb{R}^n, \mu_0; \mathbb{R}^n)$. This vector field minimises the kinetic energy $\int_{\mathbb{R}^n} |v(x)|^2 \, \mathrm{d}\mu_0(x)$ among all vector fields v satisfying the continuity equation at $t = 0$, see, e.g., [1, Section 8] for details.

It was realised by Otto [27] that the Benamou–Brenier formula can be regarded as the Riemannian distance formula associated to a formal infinite-dimensional Riemannian structure on $\mathscr{P}_2(\mathbb{R}^n)$, which can be described as follows:

- Given a smooth curve $(\mu_t)_{t\in(-\epsilon,\epsilon)}$ in $\mathscr{P}_2(\mathbb{R}^n)$ with $\mu_0 = \mu$, let $v = \nabla\psi$ be the unique (generalised) gradient vector field satisfying the continuity equation $\partial_t\mu_t + \nabla\cdot(\mu\nabla\psi) = 0$ at time 0. The velocity vector field $\nabla\psi$ is regarded as the tangent vector at $\mu \in \mathscr{P}_2(\mathbb{R}^n)$ associated to the curve $(\mu_t)_{t\in(-\epsilon,\epsilon)}$. Thus, by means of the continuity equation, the tangent space at μ is identified with the space of generalised gradients.
- Taking this identification into account, the tangent space at $\mu \in \mathscr{P}(\mathbb{R}^n)$ will be endowed with the $L^2(\mu)$-scalar product:

$$\langle \nabla\psi_1, \nabla\psi_2 \rangle_{T_\mu} := \int_{\mathbb{R}^n} \langle \nabla\psi_1(x), \nabla\psi_2(x) \rangle \, \mathrm{d}\mu(x) \, .$$

It is then clear that the Benamou–Brenier formula is precisely the formula for the Riemann distance induced by this formal infinite-dimensional Riemannian structure.

3. A gradient flow structure for reversible Markov chains

We shall now describe a (standard) discrete framework in which an analogue of Theorem 3 has been obtained.

Let $\mathcal{L} : \mathbb{R}^{\mathcal{X}} \to \mathbb{R}^{\mathcal{X}}$ be the generator of a continuous time Markov chain on a finite set $\mathcal{X}$. Thus $\mathcal{L}$ is given by

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

where $Q(x,y) \geq 0$ denotes the transition rate from x to y. We shall use the convention that $Q(x,x) = 0$ for all $x \in \mathcal{X}$. Assuming that the Markov chain is irreducible, there exists a unique invariant probability measure π on $\mathcal{X}$, which means that

$$(\mathcal{L}^*\pi)(x) := \sum_{y\in\mathcal{X}} \pi(y)Q(y,x) - \pi(x)Q(x,y) = 0 \ , \qquad x \in \mathcal{X} \ .$$

We shall assume in addition that π is reversible, i.e., the detailed balance equations

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

hold for all $x, y \in \mathcal{X}$. This assumption implies that $\mathcal{L}$ is selfadjoint as an operator on $L^2(\mathcal{X}, \pi)$. The associated Markov semigroup $(e^{t\mathcal{L}})_{t\geq 0}$ may be regarded as a discrete analogue of the heat semigroup. Note that the set of probability densities

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

is invariant under the action of the semigroup $(e^{t\mathcal{L}})_{t\geq 0}$.

Let $\mathcal{H} : \mathscr{D}(\mathcal{X}) \to \mathbb{R}$ be the relative entropy functional given by

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

It is straighforward to check that $\mathcal{H}$ decreases along solutions to the discrete "heat equation" $\partial_t \rho = \mathcal{L}\rho$. In view of Theorem 3 one might wonder whether this equation has a gradient flow structure for the relative entropy with respect to the 2-Wasserstein metric (for a suitable metric on $\mathcal{X}$). However, the discussion at the end of Section 1 demonstrates that this is not the case: since the 2-Wasserstein space does not contain any Lipschitz curves, there aren't any gradient flows at all!

However, it turns out to be possible to construct a different metric on $\mathscr{P}(\mathcal{X})$, which allows one to prove a discrete analogue of the JKO-Theorem. The construction is inspired by the Benamou–Brenier formula (2).

3.1. **Discrete transport metrics.** In order to define suitable discrete transport metrics, it will be necessary to state a discrete version of the continuity equation (1). For this purpose, let $\mathcal{E} := \{(x,y) : Q(x,y) > 0\}$ be the set of edges in the incidence graph induced by Q, and set $w(x,y) := Q(x,y)\pi(x)$. The *discrete gradient* is given by

$$\nabla : L^2(\mathcal{X}, \pi) \to L^2(\mathcal{E}, w) \ , \qquad \nabla\psi(x,y) := \psi(y) - \psi(x) \ .$$

This operator maps functions (defined on $\mathcal{X}$) to vector fields (defined on $\mathcal{E}$). The negative of its adjoint is the *discrete divergence* given by

$$\nabla\cdot : L^2(\mathcal{E}, w) \to L^2(\mathcal{X}, \pi) \ , \qquad (\nabla \cdot V)(x) := \frac{1}{2} \sum_{y\in\mathcal{X}} Q(x,y)(V(x,y) - V(y,x)) \ .$$

Note however that a problem arises if one attempts to formulate a discrete analogue of the continuity equation (1). Namely, the divergence term contains the product of the density ρ (defined on $\mathcal{X}$) and the velocity vector field V (defined on $\mathcal{E}$). Since these objects are defined on different spaces, there is no canonical way to multiply them, hence there is no canonical discretisation of the continuity equation.

In fact, there is additional freedom that we need to exploit: given a probability density $\rho \in \mathscr{D}(\mathcal{X})$, we introduce the quantity $\hat{\rho} : \mathcal{E} \to \mathbb{R}$ by

$$\hat{\rho}(x,y) := \theta(\rho(x), \rho(y)) , \tag{3}$$

where $\theta : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is a symmetric continuous function, smooth and strictly positive on $(0,\infty) \times (0,\infty)$, which will be carefully chosen below. In many applications, $\theta(r,s)$ will be a suitable mean of r and s. Having introduced this quantity, the natural discrete analogue of the continuity equation (1) is given by $\partial_t \rho + \nabla \cdot (\hat{\rho} V) = 0$, or more explicitly,

$$\partial_t \rho_t(x) + \frac{1}{2} \sum_{y \in \mathcal{X}} \big(V_t(x,y) - V_t(y,x)\big) \widehat{\rho_t}(x,y) Q(x,y) = 0 \tag{4}$$

The following definition is now a natural discrete analogue of the Benamou–Brenier formula.

Definition 4 (Discrete transport metric)**.** *For $\rho_0, \rho_1 \in \mathscr{D}(\mathcal{X})$ we define*

$$\mathcal{W}(\rho_0, \rho_1)^2 = \inf_{\rho, V} \left\{ \frac{1}{2} \int_0^1 \sum_{x,y \in \mathcal{X}} |V_t(x,y)|^2 \widehat{\rho_t}(x,y) Q(x,y) \pi(x) \, \mathrm{d}t \right\} , \tag{5}$$

where the infimum runs over all pairs $(\rho_t, V_t)_{t \in [0,1]}$ solving the continuity equation (4) *with boundary conditions $\rho|_{t=0} = \rho_0$ and $\rho|_{t=1} = \rho_1$.*

Note that the definition depends on the Markov triple $(\mathcal{X}, Q, \pi)$ as well as on the choice of the function θ. As in the continuous setting, it is possible to show that one may restrict the infimum to vector fields $(V_t)_{t \in [0,1]}$ of gradient type, i.e., one may assume that $V_t(x,y) = \nabla \psi_t(x,y)$ for some function $\psi_t : \mathcal{X} \to \mathbb{R}$. One can show that $\mathcal{W}(\rho_0, \rho_1) < \infty$ if ρ_0, ρ_1 are everywhere strictly positive. If ρ_0 or ρ_1 vanishes somewhere, then $\mathcal{W}(\rho_0, \rho_1)$ may be finite or infinite, depending on the choice of θ.

3.2. **A Riemannian structure on the space of probability measures.** The distance $\mathcal{W}$ is induced by a Riemannian structure on $\mathscr{D}_*(\mathcal{X})$, the space of strictly positive probability densities on $\mathcal{X}$. This structure is a natural discrete analogue of the one described in Section 2.1. Indeed, given a smooth curve $(\rho_t)_{t \in (-\epsilon, \epsilon)}$ in $\mathscr{D}_*(\mathcal{X})$ with $\rho_0 = \rho$, there exists a unique discrete gradient $\nabla \psi$ satisfying the discrete continuity equation

$$\partial_t \rho_t(x) + \sum_{y \in \mathcal{X}} \big(\psi_t(y) - \psi_t(x)\big) \widehat{\rho_t}(x,y) Q(x,y) = 0 \tag{6}$$

at time 0. As before, we shall regard $\nabla \psi$ as the tangent vector at ρ associated to the curve $(\rho_t)_{t \in (-\epsilon, \epsilon)}$. We thus identify the tangent space at ρ with the set of discrete gradients. Under this identification, we define a scalar product on the tangent space at $\rho \in \mathscr{P}(\mathbb{R}^n)$ by

$$\langle \nabla \psi_1, \nabla \psi_2 \rangle_{T_\rho} := \frac{1}{2} \sum_{x,y \in \mathcal{X}} \big(\psi_1(x) - \psi_1(y)\big) \big(\psi_2(x) - \psi_2(y)\big) \hat{\rho}(x,y) Q(x,y) \pi(x) .$$

The induced Riemannian distance coincides with (5), since it is not hard to show that the minimising vector field v in (5) is a discrete gradient for every $t \in (0,1)$.

3.3. **The discrete JKO-Theorem.** We are now in a position to obtain an analogue of Theorem 3 in the discrete setting. Let us first compute the gradient of the relative entropy $\mathcal{H}$ in the Riemannian structure described above. For this purpose, take a smooth curve $(\rho_t)_t$ in $\mathscr{D}_*(\mathcal{X})$ satisfying the discrete continuity equation (6). Using the detailed balance assumption, it follows that

$$\begin{aligned}
\partial_t \mathcal{H}(\rho_t) &= \sum_{x\in\mathcal{X}} \big(1+\log\rho_t(x)\big)\partial_t\rho_t(x)\pi(x)\\
&= -\sum_{x,y\in\mathcal{X}} \big(1+\log\rho_t(x)\big)\big(\psi_t(y)-\psi_t(x)\big)\widehat{\rho}_t(x,y)Q(x,y)\pi(x)\\
&= \frac{1}{2}\sum_{x,y\in\mathcal{X}} \big(\log\rho_t(x)-\log\rho_t(y)\big)\big(\psi_t(x)-\psi_t(y)\big)\widehat{\rho}_t(x,y)Q(x,y)\pi(x)\\
&= \langle \nabla\log\rho_t, \nabla\psi_t\rangle_{T_{\rho_t}}
\end{aligned}$$

On the other hand, by the definition of the gradient in the Riemannian structure inducing $\mathcal{W}$, we have $\partial_t\mathcal{H}(\rho_t) = \langle \mathrm{grad}_{\mathcal{W}}\,\mathcal{H}(\rho_t), \nabla\psi_t\rangle_{T_{\rho_t}}$. Since the computation above holds for any curve $(\rho_t)_t$ (hence for any vector field $\nabla\psi_t$), we infer that

$$\mathrm{grad}_{\mathcal{W}}\,\mathcal{H}(\rho) = \nabla\log\rho \tag{7}$$

for $\rho\in\mathscr{D}_*(\mathcal{X})$. It follows from this identity that the gradient flow equation for $\mathcal{H}$ (in the Riemannian structure associated to $\mathcal{W}$) is given by the discrete continuity equation (6) together with the additional equation $\nabla\psi_t = -\nabla\log\rho_t$. More explicitly, we arrive at

$$\partial_t\rho_t(x) - \sum_{y\in\mathcal{X}} \big(\log\rho_t(y)-\log\rho_t(x)\big)\widehat{\rho}_t(x,y)Q(x,y) = 0\ .$$

Of course, this equation depends on the choice of θ in (3), since the Riemannian metric depends on θ. Note however that this equation reduces to the discrete heat equation $\partial_t\rho = \mathcal{L}\rho$ if we choose θ in such a way that $(\log\rho(y)-\log\rho(x))\hat{\rho}(x,y) = \rho(y)-\rho(x)$. In other words, if θ is the *logarithmic mean* defined by

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

we obtain the following discrete JKO-Theorem:

Theorem 5 (Discrete JKO)**.** *The gradient flow equation for the relative entropy $\mathcal{H}$ with respect to $\mathcal{W}$ is given by the discrete heat equation $\partial_t\rho = \mathcal{L}\rho$, provided $\theta = \theta_{\log}$.*

This result has been obtained in the independent papers [22] (in the setting of Markov chains) and [24] (in the setting of reaction-diffusion systems). Related gradient flow structures for Fokker-Planck equations on graphs have been discovered in [8]. A modification of the proof above shows that the discrete porous medium equations $\partial_t\rho = \Delta\varphi(\rho)$ can be formulated as gradient flow for the entropy functional $\mathcal{F}(\rho) = \sum_{x\in\mathcal{X}} f(\rho(x))\pi(x)$ with respect to the metric $\mathcal{W}$, provided that $\theta(r,s) = \frac{\varphi(r)-\varphi(s)}{f'(r)-f'(s)}$ for some increasing function φ and some convex function f, cf. [15].

4. Discrete Entropic Ricci curvature

In view of the gradient flow result of Theorem 5, the metric $\mathcal{W}$ (with $\theta = \theta_{\log}$) can be viewed as a natural discrete counterpart of the Wasserstein metric W_2. The following definition, proposed in [22], is a discrete analogue of Definition 2 by Lott–Sturm–Villani.

Definition 6 (Discrete entropic Ricci curvature)**.** *Let $\kappa \in \mathbb{R}$. A reversible Markov chain $(\mathcal{X}, Q, \pi)$ is said to have "Ricci curvature bounded from below by κ" if every pair of probability densities $\rho_0, \rho_1 \in \mathscr{D}_*(\mathcal{X})$ can be connected by a constant speed $\mathcal{W}$-geodesic $(\rho_t)_{t\in[0,1]}$ along which the relative entropy $\mathcal{H}$ satisfies the κ-convexity inequality*

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

We shall use the notation $\mathrm{Ric}(\mathcal{X}, Q, \pi) \geq \kappa$ for brevity.

The restriction to probability densities $\rho_0, \rho_1 \in \mathscr{D}_*(\mathcal{X})$ is not essential: if $\theta = \theta_{\log}$ it can be shown that $\mathcal{W}$ is finite on the full space $\mathscr{D}(\mathcal{X})$, and the condition in Definition 6 may be imposed for all $\rho_0, \rho_1 \in \mathscr{D}(\mathcal{X})$ without changing the definition.

It has been shown in [14] that several discrete analogues of classical continuous results can be obtained from this definition. The following result is a discrete analogue of a classical result by Bakry and Émery [3].

Theorem 7 (Discrete Bakry–Émery)**.** *If $\mathrm{Ric}(\mathcal{X}, Q, \pi) \geq \kappa$ for some $\kappa > 0$, then the modified logarithmic Sobolev inequality*

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

holds for all $\rho \in \mathscr{D}_*(\mathcal{X})$, where

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

is a discrete version of the Fisher information.

There are different (non-equivalent) versions of the logarithmic Sobolev inequality that appear in discrete settings. The relevance of the inequality (MLSI(κ)) is due to the fact that it implies the exponential convergence result $\mathcal{H}(\rho_t) \leq e^{-2\kappa t}\mathcal{H}(\rho_0)$ for solutions to the discrete heat equation $\partial_t\rho = \mathcal{L}\rho$.

The following result from [14] is a discrete version of a celebrated result by Otto and Villani [28].

Theorem 8 (Discrete Otto–Villani)**.** *If $(\mathcal{X}, Q, \pi)$ satisfies (MLSI(κ)) for some $\kappa > 0$, then the modified Talagrand inequality*

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

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

In this result, $\mathbf{1}$ denotes the constant probability density corresponding to the invariant measure π. The analogous inequality with $\mathcal{W}$ replaced by W_2 has been extensively studied in continuous settings, but it is not hard to see that it can never hold in the discrete case. The inequality ($\mathrm{T}_{\mathcal{W}}(\kappa)$) provides a natural discrete substitute, which captures two different phenomena. On the one hand, ($\mathrm{T}_{\mathcal{W}}(\kappa)$) contains spectral information, since it implies the Poincaré inequality (or spectral gap inequality)

$$\|\varphi\|^2_{L^2(\mathcal{X},\pi)} \leq \frac{1}{\kappa}\mathcal{D}(\varphi) \tag{P(κ)}$$

where $\mathcal{D}(\varphi) = \frac{1}{2}\sum_{x,y\in\mathcal{X}}\big(\varphi(x)-\varphi(y)\big)^2 Q(x,y)\pi(x)$ denotes the Dirichlet energy. On the other hand, ($\mathrm{T}_{\mathcal{W}}(\kappa)$) yields the T_1-transport inequality

$$W_1(\rho,\mathbf{1}) \leq \sqrt{\frac{2}{\kappa'}\mathcal{H}(\rho)}$$

for a possibly different constant κ', where W_1 is the 1-Wasserstein distance induced by the graph distance on the incidence graph of $(\mathcal{X},Q,\pi)$. The latter inequality implies an exponential concentration inequality for π with an explicit rate, cf. [14].

4.1. **Discrete spaces with lower Ricci curvature bounds.** As discrete entropic Ricci curvature bounds have significant consequences for the Markov chains under consideration, it is of interest to obtain sharp bounds for the Ricci curvature in concrete examples. Since such bounds are defined in terms of convexity properties of the relative entropy $\mathcal{H}$, one needs to calculate the Hessian of $\mathcal{H}$ in the Riemannian structure inducing $\mathcal{W}$. Therefore, we would like to compute the second derivative of the entropy along $\mathcal{W}$-geodesics. The geodesic equations for $\mathcal{W}$ are given by the continuity equation (6) in conjunction with the equations

$$\partial_t\psi_t(x) + \frac{1}{2}\sum_{y\in\mathcal{X}}\big(\psi_t(x)-\psi_t(y)\big)^2\partial_1\theta(\rho(x),\rho(y))Q(x,y) = 0\ , \qquad x\in\mathcal{X}\ . \tag{8}$$

This equation is reminiscent of the Hamilton-Jacobi equation $\partial_t\psi + \frac{1}{2}|\nabla\psi|^2 = 0$, which appears in the description of 2-Wasserstein geodesics in continuous settings. Note however that the discrete equation (8) depends both on ρ and ψ, which is a source of additional difficulties in the discrete setting. An explicit computation based on the equations (4) and (8) shows that the second derivative of the entropy along a unit speed $\mathcal{W}$-geodesics $(\rho_t)_t$ is given by

$$\begin{aligned}
\partial_t^2\big|_{t=0}\mathcal{H}(\rho_t) &= \frac{1}{4}\sum_{x,y,z\in\mathcal{X}}\big(\psi(x)-\psi(y)\big)^2\Big(\partial_1\theta\big(\rho(x),\rho(y)\big)\big(\rho(z)-\rho(x)\big)Q(x,z) \\
&\qquad +\partial_2\theta\big(\rho(x),\rho(y)\big)\big(\rho(z)-\rho(y)\big)Q(y,z)\Big)Q(x,y)\pi(x) \\
&= -\frac{1}{2}\sum_{x,y,z\in\mathcal{X}}\Big(Q(x,z)\big(\psi(z)-\psi(x)\big) - Q(y,z)\big(\psi(z)-\psi(y)\big)\Big) \\
&\qquad \times\big(\psi(x)-\psi(y)\big)\hat\rho(x,y)Q(x,y)\pi(x) =: \mathcal{B}(\rho,\psi)\ ,
\end{aligned}$$

Proving a bound of the form $\mathrm{Ric}(\mathcal{X},Q,\pi)\geq\kappa$ thus corresponds to showing that

$$\mathcal{B}(\rho,\psi) \geq \frac{\kappa}{2}\sum_{x,y}(\psi(x)-\psi(y))^2\hat\rho(x,y)Q(x,y)\pi(x)\ . \tag{9}$$

The right-hand side above may be seen as a discrete version of the continuous formula $\int \big(\frac{1}{2}|\nabla\psi|^2\Delta\rho - \langle\nabla\Delta\psi,\nabla\psi\rangle\rho\big)\,\mathrm{d}x$. On a Riemannian manifold, such an expression can be simplified after integration by parts with the help of Bochner's identity, which asserts that

$$\frac{1}{2}\Delta\big(|\nabla\psi_t|^2\big) - \langle\nabla\psi_t,\nabla\Delta\psi_t\rangle = \frac{1}{2}|\mathrm{D}^2\psi|^2 + \mathrm{Ric}(\nabla\psi,\nabla\psi)\ .$$

Hence a lower bound on the Ricci curvature yields a corresponding bound for the Hessian of the entropy. In the discrete case there doesn't seem to be an exact analogue of Bochner's formula, and the challenge is to get around this difficulty in concrete examples of interest.

4.2. **Examples.** In recent years, discrete entropic Ricci curvature bounds have been obtained in several concrete examples.

A result by Mielke [25] asserts that every reversible Markov chain on a finite state space $\mathcal{X}$ satisfies $\mathrm{Ric}(\mathcal{X},Q,\pi)\geq\kappa$ for some (possibly negative) $\kappa\in\mathbb{R}$. Even though the Riemannian manifold $\mathscr{D}_*(\mathcal{X})$ is finite-dimensional, this result is non-trivial (and its proof is rather delicate), since the Riemannian metric inducing $\mathcal{W}$ is degenerate at the boundary of $\mathscr{D}_*(\mathcal{X})$.

If the state space $\mathcal{X}$ has a one-dimensional structure, the situation is quite well understood. Consider a birth-death chain (on a finite set $\{0,1,\ldots,N\}$ for some $N\geq 1$) with transition rates $Q(n,n+1)=a(n)$ and $Q(n,n-1)=b(n)$. In this situation, Mielke [25] obtained the following lower bounds on the Ricci curvature for birth-death processes (see also [17] for a different proof).

Theorem 9 (Ricci bounds for birth-death chains)**.** *Let $\kappa\in[0,\infty)$. Assume that the rate of birth a is non-increasing, and that the rate of death b is a non-decreasing. Assume moreover that*

$$\frac{1}{2}\big(a(n)-a(n+1)+b(n+1)-b(n)\big)+\frac{1}{2}\Theta\big(a(n)-a(n+1),b(n+1)-b(n)\big)\geq\kappa \quad (10)$$

for all $n=0,\ldots,N-1$, where

$$\Theta(\alpha,\beta)=\inf_{s,t>0}\theta(s,t)\left(\frac{\alpha}{s}+\frac{\beta}{t}\right)\,,$$

and $\theta(s,t)$ is the logarithmic mean of s and t. Then the birth-death process has Ricci curvature bounded from below by κ.

In general this result does not give the optimal constants, but the result is asymptotically sharp in the sense that one recovers the optimal Ricci curvature bounds for one-dimensional Fokker-Planck equations by passing to the continuum limit after a suitable rescaling, cf [25].

The following result, taken from [14], asserts that Ricci bounds are preserved under taking product chains. If $\mathcal{L}_i$ generates a Markov chain on $\mathcal{X}_i$ with reversible measure π_i for $i=1,2$, the corresponding product chain on $\mathcal{X}_1\times\mathcal{X}_2$ is the Markov chain with transition semigroup $e^{t\mathcal{L}}:=e^{t\mathcal{L}_1}\otimes e^{t\mathcal{L}_2}$. Note that the generator is given by

$$\mathcal{L}^{\otimes}=\mathcal{L}_1\otimes I+I\otimes\mathcal{L}_2\ .$$

We shall write $Q^{\otimes}$ accordingly. The product measure $\pi_1\otimes\pi_2$ is the reversible invariant measure for $\mathcal{L}^{\otimes}$.

Theorem 10 (Tensorisation of discrete entropic Ricci curvature)**.** *For $i = 1, 2$, let $(\mathcal{X}_i, Q_i, \pi_i)$ be reversible Markov chains with lower Ricci bounds $\kappa_i \in \mathbb{R}$. Then the associated product chain satisfies* $\mathrm{Ric}(\mathcal{X}_1 \times \mathcal{X}_2, Q^{\otimes}, \pi_1 \otimes \pi_2) \geq \min\{\kappa_1, \kappa_2\}$.

This result allows one to consider product chains in arbitrarily high dimension, without any loss in the Ricci curvature constant. As a consequence, it can be proved that the optimal Ricci curvature bound for simple random walk on the discrete cube $\{0,1\}^N$ is given by $\frac{2}{N}$. Similarly, it follows that the Ricci curvature for any regular rectangular lattice in any dimension is non-negative.

Much less is known for Markov chains in multiple dimensions which do not have a product structure, but there has been some recent progress in this direction. Using combinatorial methods, entropic Ricci curvature bounds have been obtained for the Bernoulli-Laplace model, which describes a simple random walk on a slice of the discrete cube, as well as for the random transposition model, which describes a random walk on the group of permutations [16]. A more systematic method for proving discrete Ricci bounds, inspired by the work of [6] on functional inequalities, has been developed in [17] and applied to zero-range processes on the complete graph.

We conclude by mentioning that there are several other interesting notions of discrete Ricci curvature that have been actively studied in recent years. Important examples include Ollivier's Ricci curvature [26] and the Bakry-Émery criterion [3] and its variants. It is not hard to show that the Bakry-Émery criterion can be obtained from the inequality (9) that characterizes discrete entropic Ricci curvature, by replacing the logarithmic mean $\theta_{\log}$ by the arithmetic mean. Since the proofs in [14, 16, 17] remain valid for the the arithmetic mean, these papers also yield curvature bounds in the sense of Bakry-Émery. However, there are examples in of Markov chains with non-negative entropic Ricci curvature that have strictly negative Ricci curvature in the sense of Bakry-Émery [11]. We refer to other chapters in this volume, in particular the contribution by F. Bauer, B. Hua, J. Jost, S. Liu and G. Wang for a discussion of Ollivier's Ricci curvature and the Bakry-Émery condition.

A discretisation of the Lott-Sturm-Villani theory based on the notion of approximate W_2-midpoints (different from the theory described in this note) was given by Bonciocat and Sturm [5]. Convexity of the entropy in discrete settings has been investigated along other interesting curves of probability measures in [18] and in [20].

Note added: After this short paper was written, several papers dealing with discrete entropic Ricci curvature have appeared. In particular, Erbar and Fathi [10] obtained various interesting functional inequalities under the assumption of non-negative entropic Ricci curvature. Moreover, entropic Ricci curvature bounds have been proved for several weakly interacting Markov chains (including Glauber dynamics for a class of spin systems) by Erbar, Henderson, Menz, and Tetali [12]. Related functional inequalities have been investigated by Che, Huang, Li, and Tetali [7].

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. Metric measure spaces with Riemannian Ricci curvature bounded from below. *Duke Math. J.*, 163(7):1405–1490, 2014.
[3] 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.
[4] 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.
[5] A.-I. Bonciocat and K.-Th. Sturm. Mass transportation and rough curvature bounds for discrete spaces. *J. Funct. Anal.*, 256(9):2944–2966, 2009.
[6] 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.
[7] R. Che, W. Huang, Y. Li, and P. Tetali. Convergence to global equilibrium for Fokker-Planck equations on a graph and Talagrand-type inequalities. *J. Differential Equations*, 261(4):2552–2583, 2016.
[8] S.-N. Chow, W. Huang, Y. Li, and H. Zhou. Fokker-Planck equations for a free energy functional or Markov process on a graph. *Arch. Ration. Mech. Anal.*, 203(3):969–1008, 2012.
[9] D. Cordero-Erausquin, R. J. McCann, and M. Schmuckenschläger. A Riemannian interpolation inequality à la Borell, Brascamp and Lieb. *Invent. Math.*, 146(2):219–257, 2001.
[10] M. Erbar and M. Fathi. Poincaré, modified logarithmic Sobolev and isoperimetric inequalities for Markov chains with non-negative Ricci curvature. *arXiv preprint arXiv:1612.00514*, 2016.
[11] M. Erbar, M. Fathi, J. Maas, A. Schlichting, and P. Tetali. *In preparation.*
[12] M. Erbar, C. Henderson, G. Menz, and P. Tetali. Ricci curvature bounds for weakly interacting Markov chains. *arXiv preprint arXiv:1602.05414*, 2016.
[13] M. Erbar, K. Kuwada, and K.-T. Sturm. On the equivalence of the entropic curvature-dimension condition and Bochner's inequality on metric measure spaces. *Invent. Math.*, 201(3):993–1071, 2015.
[14] M. Erbar and J. Maas. Ricci curvature of finite Markov chains via convexity of the entropy. *Arch. Ration. Mech. Anal.*, 206(3):997–1038, 2012.
[15] M. Erbar and J. Maas. Gradient flow structures for discrete porous medium equations. *Discrete Contin. Dyn. Syst.*, 34(4):1355–1374, 2014.
[16] M. Erbar, J. Maas, and P. Tetali. Discrete Ricci curvature bounds for Bernoulli-Laplace and random transposition models. *Ann. Fac. Sci. Toulouse Math. (6)*, 24(4):781–800, 2015.
[17] M. Fathi and J. Maas. Entropic Ricci curvature bounds for discrete interacting systems. *Ann. Appl. Probab.*, 26(3):1774–1806, 2016.
[18] N. Gozlan, C. Roberto, P.-M. Samson, and P. Tetali. Displacement convexity of entropy and related inequalities on graphs. *Probab. Theory Related Fields*, 160(1-2):47–94, 2014.
[19] R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998.
[20] C. Léonard. Lazy random walks and optimal transport on graphs. *Ann. Probab.*, 44(3):1864–1915, 2016.
[21] J. Lott and C. Villani. Ricci curvature for metric-measure spaces via optimal transport. *Ann. Math. (2)*, 169(3):903–991, 2009.
[22] J. Maas. Gradient flows of the entropy for finite Markov chains. *J. Funct. Anal.*, 261(8):2250–2292, 2011.
[23] R. J. McCann. A convexity principle for interacting gases. *Adv. Math.*, 128(1):153–179, 1997.
[24] A. Mielke. A gradient structure for reaction-diffusion systems and for energy-drift-diffusion systems. *Nonlinearity*, 24(4):1329–1346, 2011.
[25] A. Mielke. Geodesic convexity of the relative entropy in reversible Markov chains. *Calc. Var. Partial Differential Equations*, 48(1-2):1–31, 2013.
[26] Y. Ollivier. Ricci curvature of Markov chains on metric spaces. *J. Funct. Anal.*, 256(3):810–864, 2009.
[27] F. Otto. The geometry of dissipative evolution equations: the porous medium equation. *Comm. Partial Differential Equations*, 26(1-2):101–174, 2001.

[28] 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.

[29] 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.

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

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

Institute of Science and Technology Austria (IST Austria), Am Campus 1, 3400 Klosterneuburg, Austria

E-mail address: jan.maas@ist.ac.at