Entropic Ricci curvature on discrete spaces.

Max Fathi

IMT Toulouse

Based on works by M. Erbar, J. Maas and A. Mielke, and joint works with M. Erbar and J. Maas

April 21, 2017

Ricci curvature lower bounds on a manifold

Ricci curvature : measure of how far from being flat a manifold (M, g) is.
Spaces with constant curvature : spheres (positive curvature), $\mathbb{R}^d$ (zero curvature), hyperbolic spaces (negative curvature) : $\mathrm{Ric} = \kappa g$.
$\mathrm{Ric} \geq \kappa$: the space is "‘more curved"’ than the model space with constant curvature $\kappa \in \mathbb{R}$.

Probabilist's viewpoint: curvature lower bounds can be encoded into properties of Brownian motion on the manifold. Several interesting properties:

- exponentially fast convergence to equilibrium
- contraction of Brownian motions (in transport distances)
- functional inequalities (spectral gap, log-Sobolev,...)

Aim: get good quantitative estimates for the speed of convergence to equilibrium for high-dimensional systems. Many applications in statistical physics (particle systems), numerical analysis (MCMC simulation), statistics...

Synthetic notion of Ricci curvature lower bounds (Lott-Sturm-Villani): definition in geodesic spaces using optimal transport.

Definition

μ, ν : probability measures on M with finite second moment.

$$W_2(\mu,\nu)^2 := \inf_{\pi} \int ||x-y||_2^2 \pi(dx,dy)$$

π : coupling of μ and ν.

W_2 is a distance over $\mathcal{P}_2(M)$. Moreover, when M is a geodesic space, it is possible to build geodesics for W_2 (in $\mathcal{P}(M)$).

Entropy

Definition

Let μ be a nonnegative measure on a space X, and f a nonnegative function. We define

$$\mathrm{Ent}_\mu(f) = \int f \log f d\mu - \left(\int f d\mu\right) \log\left(\int f d\mu\right).$$

If $\nu = f\mu$ is a probability measure, we define $\mathrm{Ent}_\mu(\nu) = \mathrm{Ent}_\mu(f)$.

Convexity of the entropy

Theorem (Lott-Villani 2009, Sturm 2006)

On a Riemannian manifold, Ric $\geq \kappa$ iff the entropy (w.r.t. the volume measure) is κ-convex along W_2 geodesics, i.e.

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

We can then use this property as a *definition* of lower bounds on Ricci curvature, which makes sense on geodesic spaces.

Why?

Jordan-Kinderlehrer-Otto : The heat equation on the manifold

$$\dot{u} = \Delta u$$

can be encoded as the gradient flow ODE in $\mathcal{P}(M)$

$$\dot{u} = -\nabla \operatorname{Ent}_{Vol}(\mu)$$

with respect to the W_2 distance.
Hence convexity properties of the entropy translate into properties of the heat equation.

Markov chains

To adapt this notion to the discrete setting, we need to decide what plays the role of the distance W_2, and what measure to use instead of the volume.
Probabilist's viewpoint: just decide what plays the role of Brownian motion
Framework : continuous time Markov chains on a finite space $\mathcal{X}$.
Jumps from x to y occur with rate $K(x, y)$.
Assumption: there exists a reversible probability measure π, i.e. such that $\pi(x)K(x, y) = K(y, x)\pi(y)$. Also known as detailed balance condition. π shall play the role of the volume.

Question : is there an analogue to Ricci curvature lower bounds for Markov chains on discrete spaces?
Test case : simple random walk on the discrete hypercube (Gromov 99, Stroock 98).
Several possibilities : Bakry-Émery 1988, Ollivier 2007, Bonciocat-Sturm 2009, Erbar-Maas 2012, Mielke 2011, Gozlan-Roberto-Samson-Tetali 2014, Leonard 2013,...

To define a notion of curvature following the lSV approach, we need a geodesic distance on the space of probability measures. In the discrete setting, it turns out that W_2 is unsuitable.

To define a a suitable distance, we shall mimic the Benamou-Brenier formula for W_2:

$$W_2(\mu_0, \mu_1)^2 = \inf \int_0^1 |v_t|^2 d\mu_t$$

where the infimum is over all paths of pairs of measures and vector fields (μ_t, v_t) such that the continuity equation $\dot{\mu}_t + \mathrm{div}(v_t \mu_t) = 0$ holds.

In the discrete setting, vector fields are functions of edges $\Phi : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$.
The continuity equation then takes the form

$$\dot{\rho}_t(x) + \sum_y \Phi_t(x,y)K(x,y)\hat{\rho}_t(x,y) = 0$$

where ρ is the density with respect to π, and $\hat{\rho}$ is a nonnegative measure on $\mathcal{X} \times \mathcal{X}$ associated to ρ.
It turns out that the suitable choice is $\hat{\rho}(x,y) = \frac{\rho(x)-\rho(y)}{\log \rho(x) - \log \rho(y)}$.

Distance $\mathcal{W}$

Theorem (Maas 2011, Mielke 2012)

The evolution of the MC is a gradient flow of the entropy with respect to the distance

$$\mathcal{W}(\mu_0, \mu_1)^2 := \inf \int_0^1 \frac{1}{2} \sum_{x,y} \Phi_t(x,y)^2 \hat{\rho}_t(x,y) K(x,y) \pi(x) dt$$

where the infimum is taken over all couples of curves (ρ_t, Φ_t) satisfying

$$\dot{\rho}_t(x) + \sum_y \Phi_t(x,y) K(x,y) \hat{\rho}_t(x,y) = 0$$

with $\Phi_t : \mathcal{X} \times \mathcal{X} \longrightarrow \mathbb{R}$, ρ_t are probability densities with respect to π with $\rho_i \pi = \mu_i$, and

$$\hat{\rho}(x,y) = \frac{\rho(x) - \rho(y)}{\log \rho(x) - \log \rho(y)}.$$

Convexity of the entropy

Definition (Erbar-Maas 2012, Mielke 2012)

The Ricci curvature of the MC is bounded from below by $\kappa \in \mathbb{R}$ if for any geodesic $(\nu_t)_{0 \le t \le 1}$ in $\mathcal{P}(\mathcal{X})$, we have

$$\mathrm{Ent}_\pi(\nu_t) \le (1-t)\,\mathrm{Ent}_\pi(\nu_0) + t\,\mathrm{Ent}_\pi(\nu_1) - \frac{\kappa t(1-t)}{2}\mathcal{W}(\nu_0,\nu_1)^2.$$

Practical criterion: look at the second derivative of the entropy along geodesics, and check if it can be bounded from below.

Example 1: the simple RW on the hypercube

Simple random walk on $\{0,1\}^N$: with rate 1, choose a coordinate uniformly at random, and flip it

Theorem (Erbar-Maas, 2012)

This Markov chain has curvature bounded from below by $\frac{2}{N}$

In the limit $N \longrightarrow \infty$, after rescaling, we recover the curvature of the Gaussian space $(\mathbb{R}, \mathcal{N}(0,1))$.
Proof : direct computation in dimension 1 + tensorization.

Example 2: the simple RW on the discrete torus

Simple random walk on $\mathbb{Z}/N\mathbb{Z}$: with rate 1, jump right or left with probability $1/2$.
This MC has nonnegative curvature. Analogous situation to the BM on a torus (flat space).
Actually holds for any symmetric random walk on an abelian Cayley graph.

Example 3: ZRP on complete graph

K particles on complete graph with L sites. Jump rate $c_x(n)$ from site x.

Theorem (F.-Maas 2015)

If $c \leq c_x(n+1) - c_x(n) \leq c + \delta$ and $\delta < 2c/5$, then curvature is bounded from below by $\frac{c}{2} - \frac{5\delta}{4}$, uniformly in K and L.

Method of proof based on mimicking the Bochner identity. No general identity, but a nice inequality for chains satisfying good spatial invariance properties.

Other examples

- The exclusion process on the complete graph has positive curvature (Erbar, Maas and Tetali 2014, F. and Maas 2015);
- One-dimensional birth-death processes with monotone rates (Mielke 2012);
- Ising model and hardcore model at high temperature (Erbar, Henderson, Menz and Tetali 2016).

Functional inequalities

Theorem (Erbar-Maas, 2012)

Assuming Ric $\geq \kappa > 0$*, we have*

$$\mathrm{Var}_\pi(f) \leq \frac{1}{\kappa}\mathcal{E}(f,f) \quad \textit{(spectral gap/Poincaré inequality)}$$

$$\mathrm{Ent}_\pi(f) \leq \frac{1}{2\kappa}\mathcal{E}(f,\log f) \quad \textit{(mLSI)}$$

$$\mathrm{Ent}_\pi(f) \leq \mathcal{W}(\nu,\pi)\sqrt{\mathcal{E}(f,\log f)} - \frac{\kappa}{2}\mathcal{W}(\nu,\pi)^2 \quad \textit{(HWI).}$$

The Dirichlet form $\mathcal{E}(f,g) = \frac{1}{2}\sum(f(x)-f(y))(g(x)-g(y))K(x,y)\pi(x)$ plays the role of $\int \nabla f \cdot \nabla g dVol$.

Various applications: convergence to equilibrium, hydrodynamic limits, concentration of measure.

Other functional inequalities also hold (Cheeger isoperimetric inequality, transport-entropy, transport-information...)

The mLSI controls convergence to equilibrium in relative entropy: it holds iff

$$\mathrm{Ent}_\pi(\rho_t) \leq e^{-2\kappa t}\,\mathrm{Ent}_\pi(\rho_0).$$

where ρ_t is the density of the law at time t of the Markov chain.

Similarly, the spectral gap controls convergence to equilibrium in L^2 (the optimal constant is the smallest positive eigenvalue of the generator of the MC).

The mLSI is strictly stronger than a spectral gap.

As an application, we recover a spectral gap, a mLSI and an HWI inequality for the ZRP on the complete graph, with constants uniform in the size of the graph and of the number of particles. The spectral gap was previously obtained by Boudou, Caputo, Dai Pra and Posta, and the mLSI by Caputo, Dai Pra and Posta.

For homogeneous rates, Caputo and Posta proved the uniform mLSI in much greater generality (no restriction on δ) with a different technique, but with non-explicit constant.

The spectral gap for ZRP on the lattice, with diffusive constant (cL^{-2}) follow by comparing the Dirichlet forms.

Diameter and spectral gap

Theorem (Li-Yau 1979, Zhong-Yang 1984)

On a compact manifold with nonnegative curvature and diameter D, the spectral gap λ_1 satisfies

$$\lambda_1 \geq \frac{\pi^2}{D^2}.$$

Theorem (Erbar-F. 2016)

The same is true for Markov chains with nonnegative curvature, up to a universal constant c, and with diameter defined by

$$D := \sup \mathcal{W}(\delta_x, \delta_y).$$

Similarly, the mLSI holds, also with constant cD^{-2}.

Sketch of proof of the spectral gap

Step 1: local gradient estimates, allowing to show that given a bounded function f, $P_t f$ is $\frac{||f||_\infty}{\sqrt{2t}}$-lipschitz (with respect to $\mathcal{W}$);
Step 2: the Dirichlet form $\mathcal{E}(P_t f)$, which is the derivative of the variance along the flow, is decreasing. Hence

$$\mathrm{Var}_\pi(f) \leq 2t\mathcal{E}(f) + \mathrm{Var}_\pi(P_t f)$$

$$\leq 2t\mathcal{E}(f) + \frac{D^2}{2t}||f||_\infty^2;$$

Step 3: The HWI inequality implies

$$\mathrm{Ent}_\pi(f^2) \leq \frac{\lambda}{2}\mathcal{E}(f^2, \log f^2) + \frac{D^2}{2\lambda}\sum f(x)^2\pi(x)$$

from which we can deduce

$$\mathrm{Var}_\pi(f) \leq \frac{\lambda}{2}\mathcal{E}(f^2, \log f^2) + e^{D^2/2\lambda}\left(\sum |f(x)|\pi(x)\right)^2;$$

Step 4: Combine the two non-tight inequalities to deduce a tight Poincaré inequality.

The previous theorem cannot be expected to give good estimates for interacting particle systems. Indeed, for independent particles, the functional inequalities are dimension-free, while the squared diameter grows linearly with the number of particles.
For example, if we apply it to the ZRP with constant rates (which has nonnegative curvature) we get a mLSI with constant $\frac{c}{K^2 \log L}$
Conjecture : At fixed density $K/L \geq 1$, the sharp mLSI constant is of order $\frac{L}{K^2}$.
This may be related to the following conjecture for discrete curvature
Conjecture : In the previous theorem, we can replace the diameter bound by an "'effective gaussian diameter"' ρ for the invariant measure π such that

$$\forall A, \quad \pi(d_{\mathcal{W}}(A, x) \geq r) \leq Me^{-\rho r^2}$$

where A is any set with mass $\geq 1/2$.
The Riemannian version of this result has been established by E. Milman (2009).