

# Ricci curvature of finite Markov chains via convexity of the entropy

M. Erbar and J. Maas

Matilde Dolfato  
Mentor: Prof. Elia Brué

Università Bocconi  
Visiting Student Initiative - BIDSA

July 16, 2025

# Overview

Introduction

Background

Prerequisites

Into the paper

Our setting

Defining  $W$

Main results

Applications and other results

References

# Introduction

## Background

## Prerequisites

## Into the paper

Our setting

Defining W

Main results

## Applications and other results

## References

# What we are talking about

RICCI CURVATURE

OPTIMAL TRANSPORT

- ▶ Wasserstein geometry
- ▶ Convexity of Shannon's entropy

MARKOV CHAINS

## General theoretical framework

The heat flow is the gradient flow of the entropy **in the  $W_2$  space**

$\text{Ric} \geq \kappa \iff \kappa\text{-convexity of the entropy} \iff \kappa\text{-contractivity of the heat flow}$

## General theoretical framework

The heat flow is the gradient flow of the entropy in the  $W_2$  space

$\text{Ric} \geq \kappa \iff \kappa\text{-convexity of the entropy} \iff \kappa\text{-contractivity of the heat flow}$

Interpret the continuous time Markov chain as the heat flow  
(via semigroup)

The generator of the Markov chain equals generator of the heat flow, the Laplacian operator

## General theoretical framework

The heat flow is the gradient flow of the entropy in the  $W_2$  space

$\text{Ric} \geq \kappa \iff \kappa\text{-convexity of the entropy} \iff \kappa\text{-contractivity of the heat flow}$

Interpret the continuous time Markov chain as the heat flow  
(via semigroup)

The generator of the Markov chain equals generator of the heat flow, the Laplacian operator

- \* Define a geometry  $\mathcal{W}$  such that heat flow with a Markov kernel is the gradient flow of entropy \*

# Why do we care?

Mainly three motivations:

- ▶ obtain convergence to equilibrium of Markov chains
- ▶ study the geometry induced by Markov kernel (e.g. random walk on the hypercube, graphs)
- ▶ obtain discrete counterpart of powerful inequalities (e.g. mass concentration, Sobolev inequalities, spectral results)

## Introduction

## Background

## Prerequisites

## Into the paper

Our setting

Defining W

Main results

## Applications and other results

## References

## History on Ricci curvature and evolution of the heat flow

- ▶ McCann [1997] (displacement convexity on  $\mathbb{R}^n$ )
- ▶ Jordan-Kinderlehrer-Otto seminal work [1998]
- ▶ Otto–Villani [2000] (extension to manifold, Hessian of entropy on  $\mathcal{P}(M)$ )
- ▶ Cordero–Erausquin, McCann, Schmuckenschläger & von Renesse–Sturm [2005] (rigorous result on convexity and Ricci)
- ▶ Lott–Villani [2009] & Sturm [2006] (Ricci for metric measure spaces)
- ▶ Maas [2011] & Erbar–Maas [2012] among others (Ricci for discrete spaces)

## Introduction

## Background

## Prerequisites

### Into the paper

Our setting

Defining W

Main results

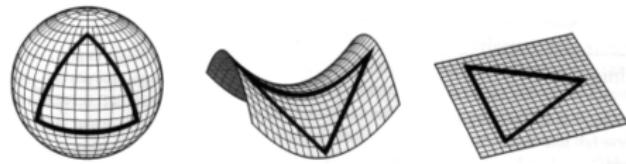
## Applications and other results

## References

# Ricci Curvature

Ricci curvature governs how volumes are distorted on the space

$$\text{Ric}_{\mathbb{R}^n} \equiv 0$$



Positive Curvature

Negative Curvature

Flat Curvature

Why a lower bound?  $(X, Y) \rightarrow \text{Ric}(X, Y) \in \mathbb{R}$

- ▶ Analytical results (functional inequalities)
- ▶ Contraction properties
- ▶ Topological obstruction

# Optimal transport I

Metric space  $(\mathcal{X}, d)$

Space of probability measures

$$\mathcal{P}(\mathcal{X}) = \{\mu : X \rightarrow \mathbb{R}_+ : \mu \text{ additive, } \mu(\mathcal{X}) = 1\}$$

- ▶ The optimal transport problem:

$$\inf_{\gamma} \left\{ \int_{X \times X} c(x, y) \, d\gamma(x, y) : \gamma \in \Gamma(\mu_0, \mu_1) \right\} \quad (\text{K})$$

for  $\mu_0, \mu_1 \in \mathcal{P}(\mathcal{X})$ ,

- ▶ Endow  $\mathcal{P}(\mathcal{X})$  with the *2-Wasserstein distance*  $W_2$ :

$$W_2^2(\mu_0, \mu_1) := \min_{\gamma} \left\{ \int_{X \times X} d^2(x, y) \, d\gamma(x, y) : \gamma \in \Gamma(\mu_0, \mu_1) \right\}$$

## Optimal transport II

- Dynamic formulation relevant for us (Benamou-Brenier):  
Given  $b$  vector field and  $\mu$  measure,

$$\mathcal{A}(b, \mu) := \int_{\mathcal{X}} |b|^2 d\mu$$

Take  $\mu_t$  curve on the space of measures  
 $\mu : [0, 1] \rightarrow \mathcal{P}(\mathcal{X}), t \mapsto \mu_t$  from  $\mu_0$  to  $\mu_1$

$$W_2^2(\mu_0, \mu_1) = \min_{b, \mu} \left\{ \int_0^1 \mathcal{A}(b_t, \mu_t) dt : \frac{d}{dt} \mu_t + \text{div}(b_t \mu_t) = 0 \right\}$$

The minimizer is a geodesic.

## Optimal transport II

- Dynamic formulation relevant for us (Benamou-Brenier):  
Given  $b$  vector field and  $\mu$  measure,

$$\mathcal{A}(b, \mu) := \int_{\mathcal{X}} |b|^2 d\mu$$

Take  $\mu_t$  curve on the space of measures  
 $\mu : [0, 1] \rightarrow \mathcal{P}(\mathcal{X}), t \mapsto \mu_t$  from  $\mu_0$  to  $\mu_1$

$$W_2^2(\mu_0, \mu_1) = \min_{b, \mu} \left\{ \int_0^1 \mathcal{A}(b_t, \mu_t) dt : \frac{d}{dt} \mu_t + \text{div}(b_t \mu_t) = 0 \right\}$$

The minimizer is a geodesic.

*Optimal transport provides the space of probability measures a geometry in the space of probability measures*

# Optimal transport III

We consider a reference measure  $\pi \in \mathcal{P}(\mathcal{X})$  and absolutely continuous measures  $\mu = \rho\pi$ ,  $\rho$  density

- ▶  $W$  convexity of the entropy:

$$\mathcal{H}(\mu_t) \leq (1-t)\mathcal{H}(\mu_0) + t\mathcal{H}(\mu_1) - \frac{\kappa}{2}t(1-t)W_2^2(\mu_0, \mu_1)$$

- ▶ Gradient flow (high-level!): given a function  $f$  with  $Dom(f) \subset \mathcal{X}$ ,  $x_t : I \rightarrow Dom(f)$  is a gradient flow for  $f$  if

$$x'_t = -\nabla f(x_t)$$

## Heat flow

Consider heat propagating on a space  $\mathcal{X}$  and a temperature function  $u(x, t)$  for  $x \in \mathcal{X}$  and  $t \in [0, T]$ ,  
the heat equation is

$$\frac{d}{dt}u(t) = \operatorname{div} \nabla u = \Delta u$$

*Heat semigroup*  $P_t := e^{t\Delta}$

## Heat flow

Consider heat propagating on a space  $\mathcal{X}$  and a temperature function  $u(x, t)$  for  $x \in \mathcal{X}$  and  $t \in [0, T]$ ,  
the heat equation is

$$\frac{d}{dt}u(t) = \operatorname{div} \nabla u = \Delta u$$

*Heat semigroup*  $P_t := e^{t\Delta}$

\* Interpret our continuous time Markov chain as the heat flow \*

Introduction

Background

Prerequisites

## Into the paper

Our setting

Defining W

Main results

Applications and other results

References

## Our setting

Finite space of states  $\mathcal{X}$ , irreducible, reversible Markov kernel  $K$

$$K(x_i, x_j) = K_{ij} = \Pr(X_t = x_j \mid X_{t-1} = x_i), \quad \sum_j K(x_i, x_j) = 1$$

Unique steady state  $\pi = \pi K$

Continuous time Markov semigroup

$P_t := e^{t(K-I)}$  Markov generator

$$\Delta := \operatorname{div} \nabla = K - I$$

Space of probability densities

$$\mathcal{P}(\mathcal{X}) := \left\{ \rho : \mathcal{X} \rightarrow \mathbb{R}_+ : \sum_{x \in \mathcal{X}} \rho(x) \pi(x) = 1 \right\}$$

## Our setting

Finite space of states  $\mathcal{X}$ , irreducible, reversible Markov kernel  $K$

$$K(x_i, x_j) = K_{ij} = \Pr(X_t = x_j \mid X_{t-1} = x_i), \quad \sum_j K(x_i, x_j) = 1$$

Unique steady state  $\pi = \pi K$

Continuous time Markov semigroup

$P_t := e^{t(K-I)}$  Markov generator

$$\Delta := \operatorname{div} \nabla = K - I$$

Space of probability densities

$$\mathcal{P}(\mathcal{X}) := \left\{ \rho : \mathcal{X} \rightarrow \mathbb{R}_+ : \sum_{x \in \mathcal{X}} \rho(x) \pi(x) = 1 \right\}$$

\* We would like to study the geometry induced by  $K$  \*

# New distance $\mathcal{W}$

Discrete action:

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

## Definition

For  $\rho_0, \rho_1 \in \mathcal{P}(\mathcal{X})$  we define

$$\mathcal{W}^2(\rho_0, \rho_1) := \inf_{\psi, \rho} \left\{ \int_0^1 \mathcal{A}(\psi_t, \rho_t) dt : \frac{d}{dt} \rho_t + \operatorname{div}(\hat{\rho}_t \nabla \psi_t) = 0 \right\}$$

with  $\rho_t, \psi_t$  sufficiently regular.

# New distance $\mathcal{W}$

Discrete action:

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

## Definition

For  $\rho_0, \rho_1 \in \mathcal{P}(\mathcal{X})$  we define

$$\mathcal{W}^2(\rho_0, \rho_1) := \inf_{\psi, \rho} \left\{ \int_0^1 \mathcal{A}(\psi_t, \rho_t) dt : \frac{d}{dt} \rho_t + \operatorname{div}(\hat{\rho}_t \nabla \psi_t) = 0 \right\}$$

with  $\rho_t, \psi_t$  sufficiently regular.

## Theorem

The heat flow  $\rho_t := P_t \rho$  is the gradient flow of Shannon's entropy  
 $\mathcal{H}(\rho) = \sum_x \rho \log \rho \pi$

$$\operatorname{grad}_{\mathcal{W}} \mathcal{H}(\rho_t) = \nabla \log(\rho_t) = -\nabla \psi_t$$

# Defining Ricci

## Definition

We say that  $K$  has a Ricci curvature bounded by below by  $\kappa \in \mathbb{R}$  if for any geodesic  $\{\rho_t\}_{t \in [0,1]}$  in  $(\mathcal{P}(\mathcal{X}), \mathcal{W})$  we have

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

and we write

$$\text{Ric}_K \geq \kappa$$

# Equivalent notions to Ricci

## Theorem

Let  $\kappa \in \mathbb{R}$ . For an irreducible and reversible Markov kernel  $(\mathcal{X}, K)$  TFAE:

- (i)  $\text{Ric}_K \geq \kappa$ ;
- (ii) For all  $\rho_0, \rho_1 \in \mathcal{P}(\mathcal{X})$  the following ‘evolution variational inequality’ holds for all  $t \geq 0$ :

$$\frac{1}{2} \frac{d^+}{dt} \mathcal{W}^2(P_t \rho_0, \rho_1) + \frac{\kappa}{2} \mathcal{W}^2(P_t \rho_0, \rho_1) \leq \mathcal{H}(\rho_1) - \mathcal{H}(P_t \rho_0);$$

- (iii) For all  $\rho \in \mathcal{P}(\mathcal{X})$   
 $\text{Hess } \mathcal{H}(\rho) \geq \kappa$ ;

- (iv) For all  $\rho_0, \rho_1 \in \mathcal{P}(\mathcal{X})$ , there exists a constant speed geodesic  $(\rho_t)_{t \in [0, T]}$  that connects them along which Shannon’s entropy is  $\kappa$ -convex

## $\kappa$ -contractivity

### Proposition

Let  $(\mathcal{X}, K)$  be as before, with  $\text{Ric}_K \geq \kappa$ . Then the continuous time Markov semigroup  $(P_t)_{t \geq 0}$  satisfies

$$\mathcal{W}(P_t \rho_0, P_t \rho_1) \leq e^{-\kappa t} \mathcal{W}(\rho_0, \rho_1)$$

for all  $\rho_0, \rho_1 \in \mathcal{P}(\mathcal{X})$  and  $t \geq 0$ .

## Introduction

## Background

## Prerequisites

## Into the paper

Our setting

Defining W

Main results

## Applications and other results

## References

## Tensorisation result

Stability of our definition of Ricci for operations on the Markov chain

$\mathcal{X} = \prod_{i=1}^n \mathcal{X}_i$ ,  $\sum_{i=1}^n \alpha_i = 1$ ,  $X = (X_1, \dots, X_n)$ , Kernel of the product chain on  $\mathcal{X}$  is  $K_\alpha$

$$K_\alpha(x, y) = \begin{cases} \sum_i \alpha_i K_i(x_i, x_i), & x_i = x_i \forall i \\ \alpha_i K_i(x_i, y_i) & x_i \neq y_i \text{ and } x_j = y_j \forall j \neq i \\ 0 & \text{otherwise} \end{cases}$$

### Theorem

Assume that  $\text{Ric}_{K_i} \geq \kappa_i$  for  $i = 1, \dots, n$ . Then we have

$$\text{Ric}_{K_\alpha} \geq \min_i \alpha_i \kappa_i$$

## Ricci of a random walk on the discrete hypercube

Let  $\mathcal{Q} = \{0, 1\}$ ,  $(\mathcal{Q}, K_{p,q})$ ,  $K(0, 1) = p, K(1, 0) = q$  for  $p, q \in (0, 1)$

Any measure  $\mu \in \mathcal{P}(\mathcal{Q})$  is  $\mu = (1 - \lambda)\delta_0 + \lambda\delta_1$  for  $\lambda \in [0, 1]$

Steady state is  $\pi$  with  $\lambda = \frac{p}{p+q}$

Let us look at  $(\mathcal{Q}^n, K_p)$  simple random walk on  $n$ -dimensional hypercube, with  $\alpha_i = 1/n$ . Maas proves

$$\text{Ric}_{K_p} \geq \kappa_p := p + p \cdot \frac{1}{2} \inf_{\lambda} \left\{ \frac{1}{\lambda(1-\lambda)} \cdot \theta(\lambda, 1-\lambda) \right\}$$

Then by tensorisation

$$\text{Ric}_{K_{p,n}} \geq \frac{1}{n} 2p$$

## References I

-  Ambrosio, L., Brué, E., Semola, D., et al. (2021).  
*Lectures on optimal transport*, volume 130.  
Springer.
-  Lee, J. M. (2003).  
*Smooth manifolds*.  
Springer.
-  Lee, J. M. (2018).  
*Introduction to Riemannian manifolds*, volume 2.  
Springer.
-  Lott, J. and Villani, C. (2009).  
Ricci curvature for metric-measure spaces via optimal transport.  
*Annals of Mathematics*, pages 903–991.

## References II

-  Maas, J. (2011).  
Gradient flows of the entropy for finite markov chains.  
*Journal of Functional Analysis*, 261(8):2250–2292.
-  Ohta, S.-I. (2014).  
*Ricci curvature, entropy, and optimal transport*, page 145–200.  
London Mathematical Society Lecture Note Series. Cambridge University Press.
-  Ollivier, Y. (2009).  
Ricci curvature of markov chains on metric spaces.  
*Journal of Functional Analysis*, 256(3):810–864.
-  Ollivier, Y. and Villani, C. (2011).  
A curved brunn–minkowski inequality on the discrete hypercube.

## References III



Villani, C. (2008).

*Optimal transport: old and new*, volume 338.  
Springer.

Thank you!



## Extra: Functional inequalities

### Theorem

Let  $K$  be an irreducible and reversible Markov kernel on a finite set  $\mathcal{X}$  with  $\text{Ric}_K \geq \kappa$  or  $\text{Ric} \geq \lambda > 0$ . The following inequalities hold:

$$\mathcal{H}(\rho) \leq \mathcal{W}(\rho, \mathbf{1}) \sqrt{\mathcal{I}(\rho)} - \frac{\kappa}{2} \mathcal{W}(\rho, \mathbf{1})^2 \quad (\text{HWI}(\kappa))$$

$$\mathcal{H}(\rho) \leq \frac{1}{2\lambda} \mathcal{I}(\rho) \quad (\text{MLSI}(\lambda))$$

$$\mathcal{W}(\rho, \mathbf{1}) \leq \sqrt{\frac{2}{\lambda} \mathcal{H}(\rho)} \quad (\text{TW}(\lambda))$$

$$\|\varphi\|_{\pi}^2 \leq \frac{1}{\lambda} \|\nabla \varphi\|_{\pi}^2 \quad (\mathsf{P}(\lambda))$$

holds for all functions  $\psi : \mathcal{X} \rightarrow \mathbb{R}$ .