Discrete Ricci curvature via convexity of the entropy

Jan Maas
University of Bonn

Joint work with Matthias Erbar

Simons Institute for the Theory of Computing
UC Berkeley

2 October 2013

Starting point

McCann '94: beautiful connection between

Starting point

McCann '94: beautiful connection between

- Boltzmann-Shannon entropy

Starting point

McCann '94: beautiful connection between

- Boltzmann-Shannon entropy
- 2-Wasserstein metric from optimal transport

Starting point

McCann '94: beautiful connection between

- Boltzmann-Shannon entropy
- 2-Wasserstein metric from optimal transport

Relative entropy

Let m be a reference measure on $\mathcal{X}$.

For $\mu \in \mathcal{P}(\mathcal{X})$: $$\mathrm{Ent}_m(\mu) = \begin{cases} \int_{\mathcal{X}} \rho(x) \log \rho(x) \, \mathrm{d}m(x), & \frac{d\mu}{dm} = \rho, \\ +\infty, & \text{otherwise.} \end{cases}$$

Optimal transport and entropy

The optimal transport problem (with quadratic cost)

Optimal transport and entropy

The optimal transport problem (with quadratic cost)
Let $(\mathcal{X}, d)$ be a Polish space and let $\mu, \nu \in \mathcal{P}(\mathcal{X})$.

Optimal transport and entropy

The optimal transport problem (with quadratic cost)
Let $(\mathcal{X}, d)$ be a Polish space and let $\mu, \nu \in \mathcal{P}(\mathcal{X})$.

Definition of the 2-Wasserstein metric

$$W_2(\mu,\nu)^2 := \inf \left\{ \int_{\mathcal{X}\times\mathcal{X}} d(x,y)^2 \, \mathrm{d}\gamma(x,y) \ : \ \gamma \text{ with marginals } \mu \text{ and } \nu \right\}$$

Optimal transport and entropy

The optimal transport problem (with quadratic cost)
Let $(\mathcal{X}, d)$ be a Polish space and let $\mu, \nu \in \mathcal{P}(\mathcal{X})$.

Definition of the 2-Wasserstein metric

$$W_2(\mu,\nu)^2 := \inf \left\{ \int_{\mathcal{X}\times\mathcal{X}} d(x,y)^2 \, \mathrm{d}\gamma(x,y) \ : \ \gamma \text{ with marginals } \mu \text{ and } \nu \right\}$$

$$= \inf \left\{ \mathbb{E}[d(X,Y)^2] \ : \ \mathrm{law}(X) = \mu, \ \mathrm{law}(Y) = \nu \right\}$$

Optimal transport and entropy

The optimal transport problem (with quadratic cost)
Let $(\mathcal{X}, d)$ be a Polish space and let $\mu, \nu \in \mathcal{P}(\mathcal{X})$.

Definition of the 2-Wasserstein metric

$$W_2(\mu,\nu)^2 := \inf\left\{ \int_{\mathcal{X}\times\mathcal{X}} d(x,y)^2 \,\mathrm{d}\gamma(x,y) \ : \ \gamma \text{ with marginals } \mu \text{ and } \nu \right\}$$
$$= \inf\left\{ \mathbb{E}[d(X,Y)^2] \ : \ \mathrm{law}(X) = \mu, \ \mathrm{law}(Y) = \nu \right\}$$

Properties

- W_2 defines a metric on $\mathcal{P}_2(\mathcal{X})$.
- $(\mathcal{X}, d)$ is a geodesic space $\Rightarrow$ $(\mathcal{P}_2(\mathcal{X}), W_2)$ is a geodesic space.

Optimal transport and entropy

Theorem (McCann '94)

The entropy is convex along geodesics in $(\mathcal{P}_2(\mathbf{R}^n), W_2)$.

Optimal transport and entropy

Theorem (McCann '94)

The entropy is convex along geodesics in $(\mathcal{P}_2(\mathbf{R}^n), W_2)$.

Construction of geodesics in the 2-Wasserstein space $\mathcal{P}_2(\mathbf{R}^n)$

Optimal transport and entropy

Theorem (McCann '94)

The entropy is convex along geodesics in $(\mathcal{P}_2(\mathbf{R}^n), W_2)$.

Construction of geodesics in the 2-Wasserstein space $\mathcal{P}_2(\mathbf{R}^n)$

- Let γ be an optimal coupling of μ and ν

Optimal transport and entropy

Theorem (McCann '94)

The entropy is convex along geodesics in $(\mathcal{P}_2(\mathbf{R}^n), W_2)$.

Construction of geodesics in the 2-Wasserstein space $\mathcal{P}_2(\mathbf{R}^n)$

- Let γ be an optimal coupling of μ and ν
- If μ is a.c., then γ is supported on the graph of a function $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ (by Brenier's Theorem).

Optimal transport and entropy

Theorem (McCann '94)

The entropy is convex along geodesics in $(\mathcal{P}_2(\mathbf{R}^n), W_2)$.

Construction of geodesics in the 2-Wasserstein space $\mathcal{P}_2(\mathbf{R}^n)$

- Let γ be an optimal coupling of μ and ν
- If μ is a.c., then γ is supported on the graph of a function $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ (by Brenier's Theorem).
- For $t \in [0,1]$, set $\Psi_t(x) := (1-t)x + t\Psi(x)$.

Optimal transport and entropy

Theorem (McCann '94)

The entropy is convex along geodesics in $(\mathcal{P}_2(\mathbf{R}^n), W_2)$.

Construction of geodesics in the 2-Wasserstein space $\mathcal{P}_2(\mathbf{R}^n)$

- Let γ be an optimal coupling of μ and ν
- If μ is a.c., then γ is supported on the graph of a function $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ (by Brenier's Theorem).
- For $t \in [0, 1]$, set $\Psi_t(x) := (1 - t)x + t\Psi(x)$.
- Let μ_t be the image measure of μ under Ψ_t.

Optimal transport and entropy

Theorem (McCann '94)

The entropy is convex along geodesics in $(\mathcal{P}_2(\mathbf{R}^n), W_2)$.

Construction of geodesics in the 2-Wasserstein space $\mathcal{P}_2(\mathbf{R}^n)$

- Let γ be an optimal coupling of μ and ν
- If μ is a.c., then γ is supported on the graph of a function $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ (by Brenier's Theorem).
- For $t \in [0, 1]$, set $\Psi_t(x) := (1 - t)x + t\Psi(x)$.
- Let μ_t be the image measure of μ under Ψ_t.
- Then: (μ_t) is a constant speed geodesic from μ to ν.

Optimal transport and entropy

Theorem (McCann '94)

The entropy is convex along geodesics in $(\mathcal{P}_2(\mathbf{R}^n), W_2)$.

Construction of geodesics in the 2-Wasserstein space $\mathcal{P}_2(\mathbf{R}^n)$

- Let γ be an optimal coupling of μ and ν
- If μ is a.c., then γ is supported on the graph of a function $\Psi : \mathbf{R}^n \to \mathbf{R}^n$ (by Brenier's Theorem).
- For $t \in [0, 1]$, set $\Psi_t(x) := (1 - t)x + t\Psi(x)$.
- Let μ_t be the image measure of μ under Ψ_t.
- Then: (μ_t) is a constant speed geodesic from μ to ν.

Highly non-linear interpolation based on *geometry* of the underlying space!

Optimal transport and entropy

Theorem (McCann '94)

The entropy is convex along geodesics in $(\mathcal{P}_2(\mathbf{R}^n), W_2)$.

Optimal transport and entropy

Theorem (McCann '94)

The entropy is convex along geodesics in $(\mathcal{P}_2(\mathbf{R}^n), W_2)$.

Theorem (Otto–Villani '00, Cordero–McCann–Schmuckenschläger '01, von Renesse–Sturm '05)

For a Riemannian manifold $\mathcal{M}$, TFAE:

1. Displacement κ-convexity of the entropy:

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

   for all 2-Wasserstein geodesics $(\mu_t)_{t\in[0,1]}$;

Optimal transport and entropy

Theorem (McCann '94)

The entropy is convex along geodesics in $(\mathcal{P}_2(\mathbf{R}^n), W_2)$.

Theorem (Otto–Villani '00, Cordero–McCann–Schmuckenschläger '01, von Renesse–Sturm '05)

For a Riemannian manifold $\mathcal{M}$, TFAE:

1. Displacement κ-convexity of the entropy:

$$\begin{aligned}\mathrm{Ent}(\mu_t) \le (1-t)\mathrm{Ent}(\mu_0) + t\mathrm{Ent}(\mu_1) \\ - \frac{\kappa}{2} t(1-t) W_2^2(\mu_0, \mu_1)\end{aligned}$$

   for all 2-Wasserstein geodesics $(\mu_t)_{t\in[0,1]}$;
2. $\mathrm{Ric} \ge \kappa$ everywhere on $\mathcal{M}$.

Optimal transport and entropy

Theorem (McCann '94)

The entropy is convex along geodesics in $(\mathcal{P}_2(\mathbf{R}^n), W_2)$.

Theorem (Otto–Villani '00, Cordero–McCann–Schmuckenschläger '01, von Renesse–Sturm '05)

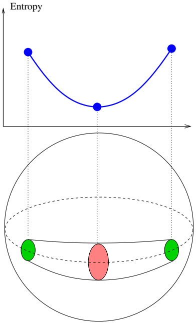
For a Riemannian manifold $\mathcal{M}$, TFAE:

1. Displacement κ-convexity of the entropy:

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

   for all 2-Wasserstein geodesics $(\mu_t)_{t\in[0,1]}$;

2. $\mathrm{Ric} \geq \kappa$ everywhere on $\mathcal{M}$.



Optimal transport and Ricci curvature II

Definition (Sturm '06, Lott–Villani '09)

A metric measure space $(\mathcal{X}, d, m)$ satisfies $\mathrm{Ric}(\mathcal{X}) \geq \kappa$ if

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

along all W_2-geodesics $(\mu_t)_{t \in [0,1]}$ in $\mathcal{P}(\mathcal{X})$.

Optimal transport and Ricci curvature II

Definition (Sturm '06, Lott–Villani '09)

A metric measure space $(\mathcal{X}, d, m)$ satisfies $\mathrm{Ric}(\mathcal{X}) \geq \kappa$ if

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

along all W_2-geodesics $(\mu_t)_{t\in[0,1]}$ in $\mathcal{P}(\mathcal{X})$.

Crucial features

Optimal transport and Ricci curvature II

Definition (STURM '06, LOTT–VILLANI '09)

A metric measure space $(\mathcal{X}, d, m)$ satisfies $\mathrm{Ric}(\mathcal{X}) \geq \kappa$ if

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

along all W_2-geodesics $(\mu_t)_{t\in[0,1]}$ in $\mathcal{P}(\mathcal{X})$.

Crucial features

- Applicable to a wide class of metric measure spaces

Optimal transport and Ricci curvature II

Definition (Sturm '06, Lott–Villani '09)

A metric measure space $(\mathcal{X}, d, m)$ satisfies $\mathrm{Ric}(\mathcal{X}) \geq \kappa$ if

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

along all W_2-geodesics $(\mu_t)_{t\in[0,1]}$ in $\mathcal{P}(\mathcal{X})$.

Crucial features

- Applicable to a wide class of metric measure spaces
- Many geometric, analytic and probabilistic consequences
  $\longrightarrow$ logarithmic Sobolev, Talagrand, Poincaré inequalities; Brunn-Minkowski.

Optimal transport and Ricci curvature II

Definition (Sturm '06, Lott–Villani '09)

A metric measure space $(\mathcal{X}, d, m)$ satisfies $\mathrm{Ric}(\mathcal{X}) \geq \kappa$ if

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

along all W_2-geodesics $(\mu_t)_{t\in[0,1]}$ in $\mathcal{P}(\mathcal{X})$.

Crucial features

- Applicable to a wide class of metric measure spaces
- Many geometric, analytic and probabilistic consequences
  $\longrightarrow$ logarithmic Sobolev, Talagrand, Poincaré inequalities; Brunn-Minkowski.
- Stability under measured Gromov–Hausdorff convergence

Optimal transport and Ricci curvature II

Definition (Sturm '06, Lott–Villani '09)

A metric measure space $(\mathcal{X}, d, m)$ satisfies $\mathrm{Ric}(\mathcal{X}) \geq \kappa$ if

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

along all W_2-geodesics $(\mu_t)_{t\in[0,1]}$ in $\mathcal{P}(\mathcal{X})$.

Crucial features

- Applicable to a wide class of metric measure spaces
- Many geometric, analytic and probabilistic consequences
  $\longrightarrow$ logarithmic Sobolev, Talagrand, Poincaré inequalities; Brunn-Minkowski.
- Stability under measured Gromov–Hausdorff convergence

But..... what about discrete spaces?

What about discrete spaces?

Example: 2-point space $\mathcal{X} = \{0, 1\}$.

What about discrete spaces?

Example: 2-point space $\mathcal{X} = \{0, 1\}$.

- Set $\mu_\alpha := (1 - \alpha)\delta_0 + \alpha\delta_1$ for $\alpha \in [0, 1]$.

What about discrete spaces?

Example: 2-point space $\mathcal{X} = \{0, 1\}$.

- Set $\mu_\alpha := (1-\alpha)\delta_0 + \alpha\delta_1$ for $\alpha \in [0,1]$. Then:

$$W_2(\mu_\alpha, \mu_\beta) = \sqrt{|\alpha - \beta|}\ .$$

What about discrete spaces?

Example: 2-point space $\mathcal{X} = \{0, 1\}$.

- Set $\mu_\alpha := (1-\alpha)\delta_0 + \alpha\delta_1$ for $\alpha \in [0,1]$. Then:

$$W_2(\mu_\alpha, \mu_\beta) = \sqrt{|\alpha - \beta|}\ .$$

- Suppose that $\big(\mu_{\alpha(t)}\big)$ is a constant speed geodesic.

What about discrete spaces?

Example: 2-point space $\mathcal{X} = \{0,1\}$.

- Set $\mu_\alpha := (1-\alpha)\delta_0 + \alpha\delta_1$ for $\alpha \in [0,1]$. Then:

$$W_2(\mu_\alpha, \mu_\beta) = \sqrt{|\alpha - \beta|}\ .$$

- Suppose that $\big(\mu_{\alpha(t)}\big)$ is a constant speed geodesic. Then:

$$W_2(\mu_{\alpha(t)}, \mu_{\alpha(s)}) = c|t-s|\ .$$

What about discrete spaces?

Example: 2-point space $\mathcal{X} = \{0, 1\}$.

- Set $\mu_\alpha := (1-\alpha)\delta_0 + \alpha\delta_1$ for $\alpha \in [0,1]$. Then:

$$W_2(\mu_\alpha, \mu_\beta) = \sqrt{|\alpha - \beta|}\ .$$

- Suppose that $\big(\mu_{\alpha(t)}\big)$ is a constant speed geodesic. Then:

$$\sqrt{|\alpha(t) - \alpha(s)|} = W_2(\mu_{\alpha(t)}, \mu_{\alpha(s)}) = c|t-s|\ .$$

What about discrete spaces?

Example: 2-point space $\mathcal{X} = \{0, 1\}$.

- Set $\mu_\alpha := (1-\alpha)\delta_0 + \alpha\delta_1$ for $\alpha \in [0,1]$. Then:

$$W_2(\mu_\alpha, \mu_\beta) = \sqrt{|\alpha - \beta|}\ .$$

- Suppose that $\left(\mu_{\alpha(t)}\right)$ is a constant speed geodesic. Then:

$$\sqrt{|\alpha(t) - \alpha(s)|} = W_2(\mu_{\alpha(t)}, \mu_{\alpha(s)}) = c|t-s|\ .$$

$\longrightarrow$ $(\alpha(t))$ is 2-Hölder, hence constant.

What about discrete spaces?

Example: 2-point space $\mathcal{X} = \{0, 1\}$.

- Set $\mu_\alpha := (1-\alpha)\delta_0 + \alpha\delta_1$ for $\alpha \in [0,1]$. Then:

$$W_2(\mu_\alpha, \mu_\beta) = \sqrt{|\alpha - \beta|}\ .$$

- Suppose that $\big(\mu_{\alpha(t)}\big)$ is a constant speed geodesic. Then:

$$\sqrt{|\alpha(t) - \alpha(s)|} = W_2(\mu_{\alpha(t)}, \mu_{\alpha(s)}) = c|t-s|\ .$$

  $\longrightarrow$ $(\alpha(t))$ is 2-Hölder, hence constant.

- Thus: there are no non-trivial W_2-geodesics. In fact:

What about discrete spaces?

Example: 2-point space $\mathcal{X} = \{0, 1\}$.

- Set $\mu_\alpha := (1-\alpha)\delta_0 + \alpha\delta_1$ for $\alpha \in [0,1]$. Then:

$$W_2(\mu_\alpha, \mu_\beta) = \sqrt{|\alpha - \beta|}\ .$$

- Suppose that $\left(\mu_{\alpha(t)}\right)$ is a constant speed geodesic. Then:

$$\sqrt{|\alpha(t) - \alpha(s)|} = W_2(\mu_{\alpha(t)}, \mu_{\alpha(s)}) = c|t-s|\ .$$

$\longrightarrow$ $(\alpha(t))$ is 2-Hölder, hence constant.

- Thus: there are no non-trivial W_2-geodesics. In fact:

$(\mathcal{P}_2(\mathcal{X}), W_2)$ is a geodesic space $\Leftrightarrow$ $(\mathcal{X}, d)$ is a geodesic space.

What about discrete spaces?

Example: 2-point space $\mathcal{X} = \{0, 1\}$.

- Set $\mu_\alpha := (1-\alpha)\delta_0 + \alpha\delta_1$ for $\alpha \in [0,1]$. Then:

$$W_2(\mu_\alpha, \mu_\beta) = \sqrt{|\alpha - \beta|}\ .$$

- Suppose that $\big(\mu_{\alpha(t)}\big)$ is a constant speed geodesic. Then:

$$\sqrt{|\alpha(t) - \alpha(s)|} = W_2(\mu_{\alpha(t)}, \mu_{\alpha(s)}) = c|t-s|\ .$$

$\longrightarrow$ $(\alpha(t))$ is 2-Hölder, hence constant.

- Thus: there are no non-trivial W_2-geodesics. In fact:

$(\mathcal{P}_2(\mathcal{X}), W_2)$ is a geodesic space $\Leftrightarrow$ $(\mathcal{X}, d)$ is a geodesic space.

Conclusion: LSV-Definition does not apply to discrete spaces.

Ricci curvature of discrete spaces

Many approaches to discrete Ricci curvature:

- W_1-contraction à la Dobrushin
  Ollivier ('09); Sammer ('05), Joulin ('09), Jost, Bauer, Hua, Liu ('11-...)
- approximate W_2-geodesics
  Bonciocat, Sturm ('09), Ollivier, Villani ('12)
- modified Bakry-Émery criterion
  Lin, Lu, S.-T. Yau ('11-...), Bauer, Horn, Lin, Lippner, Mangoubi, S.-T. Yau ('13)
- discrete displacement interpolation
  Gozlan, Roberto, Samson, Tetali ('12) [see next lecture!], Hillion ('12)

Ricci curvature of discrete spaces

Many approaches to discrete Ricci curvature:

- W_1-contraction à la Dobrushin
  Ollivier ('09); Sammer ('05), Joulin ('09), Jost, Bauer, Hua, Liu ('11-...)
- approximate W_2-geodesics
  Bonciocat, Sturm ('09), Ollivier, Villani ('12)
- modified Bakry-Émery criterion
  Lin, Lu, S.-T. Yau ('11-...), Bauer, Horn, Lin, Lippner, Mangoubi, S.-T. Yau ('13)
- discrete displacement interpolation
  Gozlan, Roberto, Samson, Tetali ('12) [see next lecture!], Hillion ('12)

Our goal: Find a notion of discrete Ricci curvature

- in the spirit of Lott-Sturm-Villani
- which allows to prove sharp functional inequalities.

Why 2-Wasserstein?

Why 2-Wasserstein?

Theorem [Jordan–Kinderlehrer–Otto '98]

The heat flow is the gradient flow of the entropy w.r.t W_2

Why 2-Wasserstein?

Theorem [Jordan–Kinderlehrer–Otto '98]

The heat flow is the gradient flow of the entropy w.r.t W_2

How to make sense of gradient flows in metric spaces?

Why 2-Wasserstein?

Theorem [Jordan–Kinderlehrer–Otto '98]

The heat flow is the gradient flow of the entropy w.r.t W_2

How to make sense of gradient flows in metric spaces?

Gradient flows in $\mathbf{R}^n$

Let $\varphi : \mathbf{R}^n \to \mathbf{R}$ smooth and convex. For $u : \mathbf{R}_+ \to \mathbf{R}^n$ TFAE:

1. u solves the gradient flow equation $u'(t) = -\nabla\varphi(u(t))$.
2. u solves the evolution variational inequality

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}|u(t) - y|^2 \leq \varphi(y) - \varphi(u(t)) \qquad \forall y \ .$$

(De Giorgi '93, Ambrosio–Gigli–Savaré '05)

Why 2-Wasserstein?

Theorem [Jordan–Kinderlehrer–Otto '98]

The heat flow is the gradient flow of the entropy w.r.t W_2 , i.e.,

$$\partial_t \mu = \Delta \mu \quad \Longleftrightarrow \quad \frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t} W_2(\mu_t, \nu)^2 \leq \mathrm{Ent}(\nu) - \mathrm{Ent}(\mu_t) \qquad \forall \nu \ .$$

How to make sense of gradient flows in metric spaces?

Gradient flows in $\mathbf{R}^n$

Let $\varphi : \mathbf{R}^n \to \mathbf{R}$ smooth and convex. For $u : \mathbf{R}_+ \to \mathbf{R}^n$ TFAE:

1. u solves the gradient flow equation $u'(t) = -\nabla\varphi(u(t))$.
2. u solves the evolution variational inequality

$$\tfrac{1}{2}\tfrac{\mathrm{d}}{\mathrm{d}t}|u(t) - y|^2 \leq \varphi(y) - \varphi(u(t)) \qquad \forall y \ .$$

(De Giorgi '93, Ambrosio–Gigli–Savaré '05)

Heat flow as gradient flow of the entropy

Many extensions have been proved:

• $\mathbf{R}^n$	Jordan–Kinderlehrer–Otto
• Riemannian manifolds	Villani, Erbar
• Hilbert spaces	Ambrosio–Savaré–Zambotti
• Finsler spaces	Ohta–Sturm
• Wiener space	Fang–Shao–Sturm
• Heisenberg group	Juillet
• Alexandrov spaces	Gigli–Kuwada–Ohta
• Metric measures spaces	Ambrosio–Gigli–Savaré

Heat flow as gradient flow of the entropy

Many extensions have been proved:

• $\mathbf{R}^n$	Jordan–Kinderlehrer–Otto
• Riemannian manifolds	Villani, Erbar
• Hilbert spaces	Ambrosio–Savaré–Zambotti
• Finsler spaces	Ohta–Sturm
• Wiener space	Fang–Shao–Sturm
• Heisenberg group	Juillet
• Alexandrov spaces	Gigli–Kuwada–Ohta
• Metric measures spaces	Ambrosio–Gigli–Savaré

Question

Is there a version of the JKO-Theorem for *discrete* spaces?

Discrete setting

Setting

- $\mathcal{X}$: finite set
- $K : \mathcal{X} \times \mathcal{X} \to \mathbf{R}_+$ Markov kernel, $\forall x \;:\; \sum_y K(x,y) = 1$
- π: reversible measure, $\forall x, y \;:\; K(x,y)\pi(x) = K(y,x)\pi(y)$

Discrete setting

Setting

- $\mathcal{X}$: finite set
- $K : \mathcal{X} \times \mathcal{X} \to \mathbf{R}_+$ Markov kernel, $\forall x \; : \; \sum_y K(x,y) = 1$
- π: reversible measure, $\forall x, y \; : \; K(x,y)\pi(x) = K(y,x)\pi(y)$

Heat flow

- Discrete Laplacian: $\Delta\psi(x) := \sum_y K(x,y)(\psi(y) - \psi(x))$
- Continuous time Markov semigroup: $P_t = e^{t\Delta}$

Discrete setting

Setting

- $\mathcal{X}$: finite set
- $K : \mathcal{X} \times \mathcal{X} \to \mathbf{R}_+$ Markov kernel, $\quad \forall x \ : \ \sum_y K(x,y) = 1$
- π: reversible measure, $\quad \forall x, y \ : \ K(x,y)\pi(x) = K(y,x)\pi(y)$

Heat flow

- Discrete Laplacian: $\Delta\psi(x) := \sum_y K(x,y)(\psi(y) - \psi(x))$
- Continuous time Markov semigroup: $P_t = e^{t\Delta}$

Relative Entropy

- $\mathcal{P}(\mathcal{X}) := \left\{ \rho : \mathcal{X} \to \mathbf{R}_+ \ \middle| \ \sum_{x \in \mathcal{X}} \rho(x)\pi(x) = 1 \right\}$,
- $\mathrm{Ent}(\rho) := \sum_{x \in \mathcal{X}} \rho(x) \log \rho(x)\, \pi(x), \qquad \rho \in \mathcal{P}(\mathcal{X})$.

Simplest non-trivial example: 2-point space

$\mathcal{X} = \{0, 1\}$ $\qquad K(0,1) = K(1,0) = 1$ $\qquad \pi(0) = \pi(1) = \frac{1}{2}$

Simplest non-trivial example: 2-point space

$\mathcal{X} = \{0, 1\}$ $\qquad$ $K(0,1) = K(1,0) = 1$ $\qquad$ $\pi(0) = \pi(1) = \frac{1}{2}$

Recall the notation: $\mu_\alpha = (1-\alpha)\delta_0 + \alpha\delta_1$.

Simplest non-trivial example: 2-point space

$\mathcal{X} = \{0, 1\}$ $\quad K(0,1) = K(1,0) = 1$ $\quad \pi(0) = \pi(1) = \frac{1}{2}$

Recall the notation: $\mu_\alpha = (1-\alpha)\delta_0 + \alpha\delta_1$.

Question

Is the heat flow on $\{0, 1\}$ the gradient flow of the entropy w.r.t. the L^2-Wasserstein metric?

Simplest non-trivial example: 2-point space

$\mathcal{X} = \{0, 1\}$ $\quad K(0,1) = K(1,0) = 1$ $\quad \pi(0) = \pi(1) = \frac{1}{2}$

Recall the notation: $\mu_\alpha = (1-\alpha)\delta_0 + \alpha\delta_1$.

Question

Is the heat flow on $\{0, 1\}$ the gradient flow of the entropy w.r.t. the L^2-Wasserstein metric?

Answer

NO! Reason: $W_2(\mu_\alpha, \mu_\beta) = \sqrt{|\alpha - \beta|}$.

Simplest non-trivial example: 2-point space

$\mathcal{X} = \{0, 1\}$ $\quad K(0,1) = K(1,0) = 1$ $\quad \pi(0) = \pi(1) = \frac{1}{2}$

Recall the notation: $\mu_\alpha = (1-\alpha)\delta_0 + \alpha\delta_1$.

Question

Is the heat flow on $\{0, 1\}$ the gradient flow of the entropy w.r.t. some other metric on $\mathcal{P}(\{-1, 1\})$?

Simplest non-trivial example: 2-point space

$\mathcal{X} = \{0, 1\}$ $\quad K(0,1) = K(1,0) = 1$ $\quad \pi(0) = \pi(1) = \frac{1}{2}$

Recall the notation: $\mu_\alpha = (1 - \alpha)\delta_0 + \alpha\delta_1$.

Question

Is the heat flow on $\{0, 1\}$ the gradient flow of the entropy w.r.t. some other metric on $\mathcal{P}(\{-1, 1\})$?

Answer

YES!

Simplest non-trivial example: 2-point space

$\mathcal{X} = \{0, 1\}$ $\quad K(0,1) = K(1,0) = 1$ $\quad \pi(0) = \pi(1) = \frac{1}{2}$

Recall the notation: $\mu_\alpha = (1-\alpha)\delta_0 + \alpha\delta_1$.

Question

Is the heat flow on $\{0,1\}$ the gradient flow of the entropy w.r.t. some other metric on $\mathcal{P}(\{-1,1\})$?

Answer

YES!

Proposition [M. 2011]

The heat flow is the gradient flow of Ent w.r.t. the metric $\mathcal{W}$, where

$$\mathcal{W}(\mu_\alpha, \mu_\beta) := \sqrt{2}\int_\alpha^\beta \sqrt{\frac{\mathrm{arctanh}(2r-1)}{2r-1}}\,\mathrm{d}r, \qquad 0 \le \alpha \le \beta \le 1.$$

Simplest non-trivial example: 2-point space

$\mathcal{X} = \{0, 1\}$ $\quad K(0,1) = K(1,0) = 1$ $\quad \pi(0) = \pi(1) = \frac{1}{2}$

Recall the notation: $\mu_\alpha = (1-\alpha)\delta_0 + \alpha\delta_1$.

Question

Is the heat flow on $\{0, 1\}$ the gradient flow of the entropy w.r.t. some other metric on $\mathcal{P}(\{-1, 1\})$?

Answer

YES!

Proposition [M. 2011]

The heat flow is the gradient flow of Ent w.r.t. the metric $\mathcal{W}$, where

$$\mathcal{W}(\mu_\alpha, \mu_\beta) := \sqrt{2}\int_\alpha^\beta \sqrt{\frac{\operatorname{arctanh}(2r-1)}{2r-1}}\,\mathrm{d}r, \qquad 0 \le \alpha \le \beta \le 1.$$

How to generalise this to the general discrete case?

Back to $\mathbf{R}^n$: W_2 as Riemannian metric (Otto '01)

- If $t \mapsto \rho_t$ is smooth, the continuity equation

$$\partial_t \rho + \mathrm{div}(\rho \Psi) = 0$$

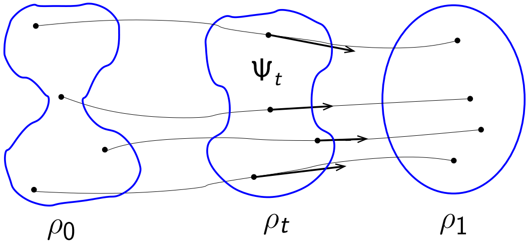
holds for some velocity vector field $\Psi(t, x)$.

Back to $\mathbf{R}^n$: W_2 as Riemannian metric (Otto '01)

- If $t \mapsto \rho_t$ is smooth, the continuity equation

$$\partial_t \rho + \mathrm{div}(\rho \Psi) = 0$$

holds for some velocity vector field $\Psi(t, x)$.

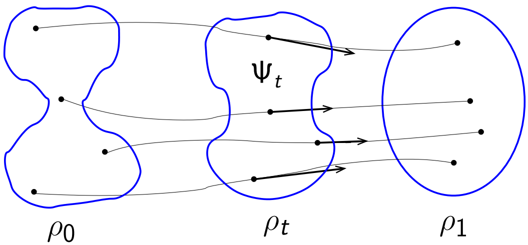


Back to $\mathbf{R}^n$: W_2 as Riemannian metric (Otto '01)

- If $t \mapsto \rho_t$ is smooth, the continuity equation

$$\partial_t \rho + \mathrm{div}(\rho \Psi) = 0$$

  holds for some velocity vector field $\Psi(t, x)$.
- For a.e. t, $\exists$ a unique gradient $\Psi_t = \nabla \psi_t$ solving cont.eq.

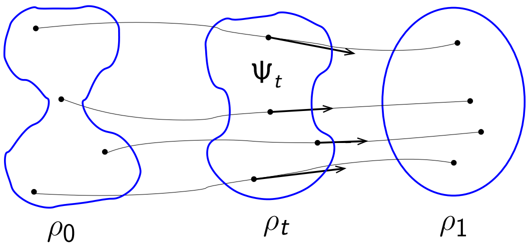


Back to $\mathbf{R}^n$: W_2 as Riemannian metric (Otto '01)

- If $t \mapsto \rho_t$ is smooth, the continuity equation

$$\partial_t \rho + \mathrm{div}(\rho \Psi) = 0$$

  holds for some velocity vector field $\Psi(t, x)$.
- For a.e. t, $\exists$ a unique gradient $\Psi_t = \nabla \psi_t$ solving cont.eq.
- Regard $\nabla \psi_t$ as "tangent vector" at ρ_t.



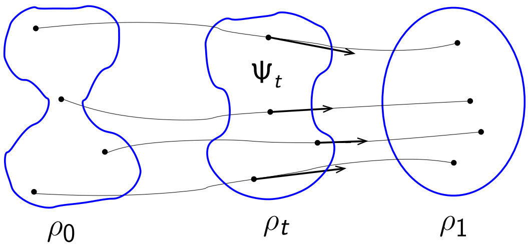
Back to $\mathbf{R}^n$: W_2 as Riemannian metric (Otto '01)

- If $t \mapsto \rho_t$ is smooth, the continuity equation

$$\partial_t \rho + \mathrm{div}(\rho \Psi) = 0$$

  holds for some velocity vector field $\Psi(t, x)$.
- For a.e. t, $\exists$ a unique gradient $\Psi_t = \nabla \psi_t$ solving cont.eq.
- Regard $\nabla \psi_t$ as "tangent vector" at ρ_t.
- For $\rho \in \mathcal{P}(\mathbf{R}^n)$, define an inner product by

$$\langle \nabla \varphi, \nabla \psi \rangle_\rho = \int_{\mathbf{R}^n} \langle \nabla \varphi(x), \nabla \psi(x) \rangle \rho(x) dx \; .$$



Back to $\mathbf{R}^n$: W_2 as Riemannian metric (Otto '01)

- If $t \mapsto \rho_t$ is smooth, the continuity equation
$$\partial_t \rho + \mathrm{div}(\rho \Psi) = 0$$
holds for some velocity vector field $\Psi(t,x)$.
- For a.e. t, $\exists$ a unique gradient $\Psi_t = \nabla \psi_t$ solving cont.eq.
- Regard $\nabla \psi_t$ as "tangent vector" at ρ_t.
- For $\rho \in \mathcal{P}(\mathbf{R}^n)$, define an inner product by
$$\langle \nabla \varphi, \nabla \psi \rangle_\rho = \int_{\mathbf{R}^n} \langle \nabla \varphi(x), \nabla \psi(x) \rangle \rho(x) dx \, .$$

Associated Riemannian distance (Benamou-Brenier formula)

$$\inf_{\rho_\cdot, \psi_\cdot} \left\{ \int_0^1 \|\nabla \psi_t\|_{\rho_t}^2 \, \mathrm{d}t \ : \partial_t \rho + \mathrm{div}(\rho \nabla \psi) = 0 \, , \right.$$
$$\left. \rho_{t=0} = \rho_0 \, , \quad \rho_{t=1} = \rho_1 \right\} .$$

Back to $\mathbf{R}^n$: W_2 as Riemannian metric (Otto '01)

- If $t \mapsto \rho_t$ is smooth, the continuity equation
$$\partial_t \rho + \mathrm{div}(\rho \Psi) = 0$$
holds for some velocity vector field $\Psi(t, x)$.
- For a.e. t, $\exists$ a unique gradient $\Psi_t = \nabla \psi_t$ solving cont.eq.
- Regard $\nabla \psi_t$ as "tangent vector" at ρ_t.
- For $\rho \in \mathcal{P}(\mathbf{R}^n)$, define an inner product by
$$\langle \nabla \varphi, \nabla \psi \rangle_\rho = \int_{\mathbf{R}^n} \langle \nabla \varphi(x), \nabla \psi(x) \rangle \rho(x) dx \ .$$

Associated Riemannian distance (Benamou-Brenier formula)

$$W_2(\rho_0, \rho_1)^2 = \inf_{\rho_\cdot, \psi_\cdot} \left\{ \int_0^1 \|\nabla \psi_t\|_{\rho_t}^2 \, \mathrm{d}t \ : \partial_t \rho + \mathrm{div}(\rho \nabla \psi) = 0 \ , \right.$$
$$\left. \rho_{t=0} = \rho_0 \ , \quad \rho_{t=1} = \rho_1 \right\} .$$

Definition of the metric $\mathcal{W}$

Benamou-Brenier formula in $\mathbf{R}^n$

$$W_2^2(\rho_0, \rho_1) = \inf_{\rho,\psi} \left\{ \int_0^1 \int_{\mathbf{R}^n} |\nabla \psi_t(x)|^2 \, \rho_t(x) \, \mathrm{d}x \, \mathrm{d}t \right\}$$

$$\text{s.t.} \quad \partial_t \rho + \mathrm{div}(\rho \nabla \psi) = 0 \text{ and } \rho_{t=0} = \rho_0, \; \rho_{t=1} = \rho_1 \; .$$

Definition of the metric $\mathcal{W}$

Benamou-Brenier formula in $\mathbf{R}^n$

$$W_2^2(\rho_0, \rho_1) = \inf_{\rho,\psi} \left\{ \int_0^1 \int_{\mathbf{R}^n} |\nabla \psi_t(x)|^2 \, \rho_t(x) \, \mathrm{d}x \, \mathrm{d}t \right\}$$

$$\text{s.t.} \quad \partial_t \rho + \mathrm{div}(\rho \nabla \psi) = 0 \text{ and } \rho_{t=0} = \rho_0, \ \rho_{t=1} = \rho_1 \ .$$

Definition in the discrete case (M. 2011)

$$\mathcal{W}(\rho_0, \rho_1)^2$$

$$:= \inf_{\rho,\psi} \left\{ \qquad \right\}$$

s.t.

Definition of the metric $\mathcal{W}$

Benamou-Brenier formula in $\mathbf{R}^n$

$$W_2^2(\rho_0, \rho_1) = \inf_{\rho,\psi} \left\{ \int_0^1 \int_{\mathbf{R}^n} |\nabla \psi_t(x)|^2 \, \rho_t(x) \, \mathrm{d}x \, \mathrm{d}t \right\}$$

$$\text{s.t.} \quad \partial_t \rho + \mathrm{div}(\rho \nabla \psi) = 0 \text{ and } \rho_{t=0} = \rho_0, \ \rho_{t=1} = \rho_1 \ .$$

Definition in the discrete case (M. 2011)

$$\mathcal{W}(\rho_0, \rho_1)^2 := \inf_{\rho,\psi} \left\{ \int_0^1 \qquad \mathrm{d}t \right\}$$

s.t.

Definition of the metric $\mathcal{W}$

Benamou-Brenier formula in $\mathbf{R}^n$

$$W_2^2(\rho_0, \rho_1) = \inf_{\rho,\psi} \left\{ \int_0^1 \int_{\mathbf{R}^n} |\nabla \psi_t(x)|^2 \, \rho_t(x) \, \mathrm{d}x \, \mathrm{d}t \right\}$$

$$\text{s.t.} \quad \partial_t \rho + \mathrm{div}(\rho \nabla \psi) = 0 \text{ and } \rho_{t=0} = \rho_0, \ \rho_{t=1} = \rho_1 \ .$$

Definition in the discrete case (M. 2011)

$$\mathcal{W}(\rho_0, \rho_1)^2$$

$$:= \inf_{\rho,\psi} \left\{ \frac{1}{2} \int_0^1 \sum_{x,y \in \mathcal{X}} \qquad K(x,y)\pi(x) \, \mathrm{d}t \right\}$$

s.t.

Definition of the metric $\mathcal{W}$

Benamou-Brenier formula in $\mathbf{R}^n$

$$W_2^2(\rho_0, \rho_1) = \inf_{\rho,\psi} \left\{ \int_0^1 \int_{\mathbf{R}^n} |\nabla \psi_t(x)|^2 \, \rho_t(x) \, \mathrm{d}x \, \mathrm{d}t \right\}$$

$$\text{s.t.} \quad \partial_t \rho + \mathrm{div}(\rho \nabla \psi) = 0 \text{ and } \rho_{t=0} = \rho_0, \ \rho_{t=1} = \rho_1 \ .$$

Definition in the discrete case (M. 2011)

$$\mathcal{W}(\rho_0, \rho_1)^2$$

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

s.t.

Definition of the metric $\mathcal{W}$

Benamou-Brenier formula in $\mathbf{R}^n$

$$W_2^2(\rho_0, \rho_1) = \inf_{\rho,\psi} \left\{ \int_0^1 \int_{\mathbf{R}^n} |\nabla \psi_t(x)|^2 \rho_t(x) \, \mathrm{d}x \, \mathrm{d}t \right\}$$

$$\text{s.t.} \quad \partial_t \rho + \mathrm{div}(\rho \nabla \psi) = 0 \text{ and } \rho_{t=0} = \rho_0, \; \rho_{t=1} = \rho_1 \; .$$

Definition in the discrete case (M. 2011)

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

s.t.

Problem: ρ is defined on vertices, $\nabla \psi$ is defined on edges

$\rightsquigarrow$ No canonical way to define the product $\rho \cdot \nabla \psi$!

Definition of the metric $\mathcal{W}$

Benamou-Brenier formula in $\mathbf{R}^n$

$$W_2^2(\rho_0, \rho_1) = \inf_{\rho,\psi} \left\{ \int_0^1 \int_{\mathbf{R}^n} |\nabla \psi_t(x)|^2 \, \rho_t(x) \, \mathrm{d}x \, \mathrm{d}t \right\}$$

$$\text{s.t.} \quad \partial_t \rho + \mathrm{div}(\rho \nabla \psi) = 0 \text{ and } \rho_{t=0} = \rho_0, \; \rho_{t=1} = \rho_1 \; .$$

Definition in the discrete case (M. 2011)

$$\mathcal{W}(\rho_0, \rho_1)^2$$

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

s.t.

Use the *logarithmic mean* as the "density on an edge"!

$$\hat{\rho}(x,y) := \int_0^1 \rho(x)^{1-\alpha} \rho(y)^\alpha \, \mathrm{d}\alpha = \frac{\rho(x) - \rho(y)}{\log \rho(x) - \log \rho(y)}$$

Definition of the metric $\mathcal{W}$

Benamou-Brenier formula in $\mathbf{R}^n$

$$W_2^2(\rho_0, \rho_1) = \inf_{\rho,\psi} \left\{ \int_0^1 \int_{\mathbf{R}^n} |\nabla \psi_t(x)|^2 \rho_t(x) \, \mathrm{d}x \, \mathrm{d}t \right\}$$

$$\text{s.t.} \quad \partial_t \rho + \mathrm{div}(\rho \nabla \psi) = 0 \text{ and } \rho_{t=0} = \rho_0, \; \rho_{t=1} = \rho_1 \; .$$

Definition in the discrete case (M. 2011)

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

s.t.

Use the *logarithmic mean* as the “density on an edge”!

$$\hat{\rho}(x,y) := \int_0^1 \rho(x)^{1-\alpha} \rho(y)^{\alpha} \, \mathrm{d}\alpha = \frac{\rho(x) - \rho(y)}{\log \rho(x) - \log \rho(y)}$$

Definition of the metric $\mathcal{W}$

Benamou-Brenier formula in $\mathbf{R}^n$

$$W_2^2(\rho_0,\rho_1) = \inf_{\rho,\psi}\left\{ \int_0^1 \int_{\mathbf{R}^n} |\nabla\psi_t(x)|^2 \rho_t(x)\,\mathrm{d}x\,\mathrm{d}t \right\}$$

$$\text{s.t.} \quad \partial_t\rho + \mathrm{div}(\rho\nabla\psi) = 0 \text{ and } \rho_{t=0} = \rho_0,\ \rho_{t=1} = \rho_1 \ .$$

Definition in the discrete case (M. 2011)

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

$$\text{s.t.} \quad \frac{\mathrm{d}}{\mathrm{d}t}\rho_t(x) + \sum_{y\in\mathcal{X}} \hat\rho_t(x,y)(\psi_t(x)-\psi_t(y))K(x,y) = 0 \qquad \forall x$$

Use the *logarithmic mean* as the “density on an edge”!

$$\hat\rho(x,y) := \int_0^1 \rho(x)^{1-\alpha}\rho(y)^\alpha\,\mathrm{d}\alpha = \frac{\rho(x)-\rho(y)}{\log\rho(x)-\log\rho(y)}$$

Results

Basic properties of $\mathcal{W}$ (M. '11)

- $\mathcal{W}$ defines a metric on $\mathcal{P}(\mathcal{X})$.

Results

Basic properties of $\mathcal{W}$ (M. '11)

- $\mathcal{W}$ defines a metric on $\mathcal{P}(\mathcal{X})$.
- $\mathcal{W}$ is induced by a Riemannian structure on $\mathcal{P}_{>0}(\mathcal{X})$.

Results

Basic properties of $\mathcal{W}$ (M. '11)

- $\mathcal{W}$ defines a metric on $\mathcal{P}(\mathcal{X})$.
- $\mathcal{W}$ is induced by a Riemannian structure on $\mathcal{P}_{>0}(\mathcal{X})$.
- The tangent space at ρ is the set of discrete gradients with

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

Results

Basic properties of $\mathcal{W}$ (M. '11)

- $\mathcal{W}$ defines a metric on $\mathcal{P}(\mathcal{X})$.
- $\mathcal{W}$ is induced by a Riemannian structure on $\mathcal{P}_{>0}(\mathcal{X})$.
- The tangent space at ρ is the set of discrete gradients with

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

Theorem [Discrete JKO] (M. '11)

The heat flow is the gradient flow of the entropy w.r.t. $\mathcal{W}$.

Results

Basic properties of $\mathcal{W}$ (M. '11)

- $\mathcal{W}$ defines a metric on $\mathcal{P}(\mathcal{X})$.
- $\mathcal{W}$ is induced by a Riemannian structure on $\mathcal{P}_{>0}(\mathcal{X})$.
- The tangent space at ρ is the set of discrete gradients with

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

Theorem [Discrete JKO] (M. '11)

The heat flow is the gradient flow of the entropy w.r.t. $\mathcal{W}$.

Independent works by Chow–Huang–Li–Zhou '12 and Mielke '11.

Why the logarithmic mean?

Formal proof of the JKO-Theorem

1. If (ρ_t, ψ_t) satisfy the cont. eq. $\partial_t \rho + \mathsf{div}(\rho \nabla \psi) = 0$, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Ent}(\rho_t) = -\langle \log \rho_t, \mathsf{div}(\rho_t \nabla \psi_t) \rangle = \langle \nabla \log \rho_t, \rho_t \nabla \psi_t \rangle \ .$$

$\longrightarrow \quad \mathsf{grad}_{W_2} \mathrm{Ent}(\rho) = \nabla \log \rho$

Why the logarithmic mean?

Formal proof of the JKO-Theorem

1. If (ρ_t, ψ_t) satisfy the cont. eq. $\partial_t \rho + \mathrm{div}(\rho \nabla \psi) = 0$, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Ent}(\rho_t) = -\langle \log \rho_t, \mathrm{div}(\rho_t \nabla \psi_t) \rangle = \langle \nabla \log \rho_t, \rho_t \nabla \psi_t \rangle \ .$$

$\longrightarrow \quad \mathrm{grad}_{W_2} \mathrm{Ent}(\rho) = \nabla \log \rho$

2. If ρ_t solves the heat equation in $\mathbf{R}^n$, then

$$\partial_t \rho = \mathrm{div}(\nabla \rho) = -\,\mathrm{div}(\rho \nabla \psi) \ .$$

provided $\psi = -\log \rho$.

$\longrightarrow$ Tangent vector along the heat flow is $-\nabla \log \rho$.

Why the logarithmic mean?

Formal proof of the JKO-Theorem

1. If (ρ_t, ψ_t) satisfy the cont. eq. $\partial_t \rho + \mathrm{div}(\rho \nabla \psi) = 0$, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{Ent}(\rho_t) = -\langle \log \rho_t, \mathrm{div}(\rho_t \nabla \psi_t)\rangle = \langle \nabla \log \rho_t, \rho_t \nabla \psi_t \rangle \ .$$

$\longrightarrow \quad \mathrm{grad}_{W_2} \mathrm{Ent}(\rho) = \nabla \log \rho$

2. If ρ_t solves the heat equation in $\mathbf{R}^n$, then

$$\partial_t \rho = \mathrm{div}(\nabla \rho) = -\,\mathrm{div}(\rho \nabla \psi) \ .$$

provided $\psi = -\log \rho$.

$\longrightarrow \quad$ Tangent vector along the heat flow is $-\nabla \log \rho$.

Logarithmic mean compensates for the lack of a discrete chain rule:

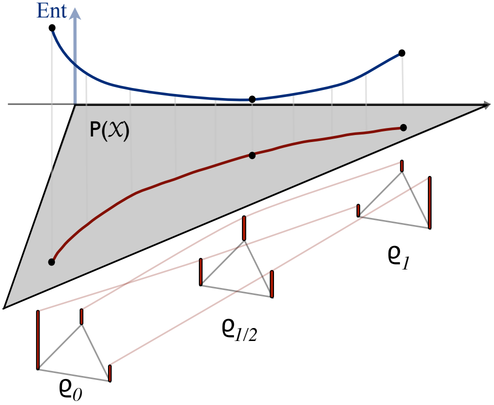
$$\rho(x) - \rho(y) = \hat{\rho}(x, y)\big(\log \rho(x) - \log \rho(y)\big)$$

Ricci curvature of Markov chains

Discrete analogue of Lott–Sturm–Villani:

Definition (Erbar, M. 2012)

We say that $(\mathcal{X}, K, \pi)$ has Ricci curvature bounded from below by $\kappa \in \mathbf{R}$ if the entropy is κ-convex along geodesics in $(\mathcal{P}(\mathcal{X}), \mathcal{W})$.



Consequences: Sharp functional inequalities I

- Let $(\mathcal{X}, K, \pi)$ be a reversible Markov chain.

Consequences: Sharp functional inequalities I

- Let $(\mathcal{X}, K, \pi)$ be a reversible Markov chain.
- Discrete analogue of the Fisher information

$$\mathcal{I}(\rho) = \frac{1}{2} \sum_{x,y} \left(\rho(x) - \rho(y)\right) \left(\log \rho(x) - \log \rho(y)\right) K(x,y) \pi(x)$$

Consequences: Sharp functional inequalities I

- Let $(\mathcal{X}, K, \pi)$ be a reversible Markov chain.
- Discrete analogue of the Fisher information

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

Discrete Bakry-Émery Theorem (Erbar, M. '12)

If $\mathrm{Ric}(K) \geq \kappa > 0$, then the modified log-Sobolev inequality holds:

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

for all $\rho \in \mathcal{P}_{>0}(\mathcal{X})$.

Consequences: Sharp functional inequalities I

- Let $(\mathcal{X}, K, \pi)$ be a reversible Markov chain.
- Discrete analogue of the Fisher information

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

Discrete Bakry-Émery Theorem (Erbar, M. '12)

If $\mathrm{Ric}(K) \geq \kappa > 0$, then the modified log-Sobolev inequality holds:

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

for all $\rho \in \mathcal{P}_{>0}(\mathcal{X})$.

- $\mathrm{mLSI}(\kappa)$ has been extensively studied in discrete settings (Bobkov–Tetali '06, ...)

Consequences: Sharp functional inequalities I

- Let $(\mathcal{X}, K, \pi)$ be a reversible Markov chain.
- Discrete analogue of the Fisher information

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

Discrete Bakry-Émery Theorem (Erbar, M. '12)

If $\mathrm{Ric}(K) \geq \kappa > 0$, then the modified log-Sobolev inequality holds:

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

for all $\rho \in \mathcal{P}_{>0}(\mathcal{X})$.

- $\mathrm{mLSI}(\kappa)$ has been extensively studied in discrete settings (Bobkov–Tetali '06, ...)
- $\mathrm{mLSI}(\kappa)$ is equivalent to the exponential decay estimate

$$\mathrm{Ent}(P_t\rho) \leq e^{-2\kappa t}\,\mathrm{Ent}(\rho).$$

Consequences: Sharp functional inequalities II

Let $(\mathcal{X}, K, \pi)$ be a reversible Markov chain. Let $\kappa > 0$.

Discrete Otto-Villani Theorem (Erbar, M. '12)

If mLSI(κ) holds, then the modified Talagrand inequality holds, i.e.,

$$\mathcal{W}(\rho, \mathbf{1})^2 \leq \frac{2}{\kappa} \operatorname{Ent}(\rho) . \qquad (\mathsf{mTal}(\kappa))$$

Consequences: Sharp functional inequalities II

Let $(\mathcal{X}, K, \pi)$ be a reversible Markov chain. Let $\kappa > 0$.

Discrete Otto-Villani Theorem (Erbar, M. '12)

If mLSI(κ) holds, then the modified Talagrand inequality holds, i.e.,

$$\mathcal{W}(\rho, \mathbf{1})^2 \leq \frac{2}{\kappa} \mathrm{Ent}(\rho) . \qquad (\mathrm{mTal}(\kappa))$$

The analogous inequality with W_2 *never* holds in discrete settings!

Consequences: Sharp functional inequalities II

Let $(\mathcal{X}, K, \pi)$ be a reversible Markov chain. Let $\kappa > 0$.

Discrete Otto-Villani Theorem (Erbar, M. '12)

If mLSI(κ) holds, then the modified Talagrand inequality holds, i.e.,

$$\mathcal{W}(\rho, \mathbf{1})^2 \leq \frac{2}{\kappa} \mathrm{Ent}(\rho) \,. \qquad (\mathrm{mTal}(\kappa))$$

The analogous inequality with W_2 *never* holds in discrete settings!

If mTal(κ) holds, then

- the Poincaré inequality holds:

$$\sum_x \psi(x)^2 \pi(x) \leq \frac{1}{2\kappa} \sum_{x,y} (\psi(x) - \psi(y))^2 K(x,y) \pi(x)$$

  whenever $\sum_x \psi(x)\pi(x) = 0$.

Consequences: Sharp functional inequalities II

Let $(\mathcal{X}, K, \pi)$ be a reversible Markov chain. Let $\kappa > 0$.

Discrete Otto-Villani Theorem (Erbar, M. '12)

If mLSI(κ) holds, then the modified Talagrand inequality holds, i.e.,

$$\mathcal{W}(\rho, \mathbf{1})^2 \le \frac{2}{\kappa} \operatorname{Ent}(\rho) . \qquad (\mathrm{mTal}(\kappa))$$

The analogous inequality with W_2 *never* holds in discrete settings!

If mTal(κ) holds, then

- the Poincaré inequality holds:

$$\sum_x \psi(x)^2 \pi(x) \le \frac{1}{2\kappa} \sum_{x,y} (\psi(x) - \psi(y))^2 K(x,y) \pi(x)$$

  whenever $\sum_x \psi(x)\pi(x) = 0$.
- the T_1-inequality holds: $W_1(\rho, \mathbf{1})^2 \le \frac{1}{\kappa} \operatorname{Ent}(\rho)$.

Ricci bounds: examples

Ricci bounds: examples

Theorem (Mielke 2012)

- For every finite reversible Markov chain: $\exists \kappa \in \mathbf{R}$ such that $\mathrm{Ric}(K) \geq \kappa$.

Ricci bounds: examples

Theorem (Mielke 2012)

- For every finite reversible Markov chain: $\exists \kappa \in \mathbf{R}$ such that $\mathrm{Ric}(K) \geq \kappa$.
- Finite volume discretisations of Fokker-Planck equations in 1D.

Ricci bounds: examples

Theorem (Mielke 2012)

- For every finite reversible Markov chain: $\exists \kappa \in \mathbf{R}$ such that $\mathrm{Ric}(K) \geq \kappa$.
- Finite volume discretisations of Fokker-Planck equations in 1D.

Theorem (Erbar, M. 2012)

Let $(\mathcal{X}_i, K_i, \pi_i)$ be reversible finite Markov chains and let $(\mathcal{X}, K, \pi)$ be the product chain. Then:

$$\mathrm{Ric}(\mathcal{X}_i, K_i, \pi_i) \geq \kappa_i \qquad \Longrightarrow \qquad \mathrm{Ric}(\mathcal{X}, K, \pi) \geq \frac{1}{n} \min_i \kappa_i$$

Ricci bounds: examples

Theorem (Mielke 2012)

- For every finite reversible Markov chain: $\exists \kappa \in \mathbf{R}$ such that $\mathrm{Ric}(K) \geq \kappa$.
- Finite volume discretisations of Fokker-Planck equations in 1D.

Theorem (Erbar, M. 2012)

Let $(\mathcal{X}_i, K_i, \pi_i)$ be reversible finite Markov chains and let $(\mathcal{X}, K, \pi)$ be the product chain. Then:

$$\mathrm{Ric}(\mathcal{X}_i, K_i, \pi_i) \geq \kappa_i \qquad \Longrightarrow \qquad \mathrm{Ric}(\mathcal{X}, K, \pi) \geq \frac{1}{n} \min_i \kappa_i$$

- Dimension-independent bounds

Ricci bounds: examples

Theorem (Mielke 2012)

- For every finite reversible Markov chain: $\exists \kappa \in \mathbf{R}$ such that $\mathrm{Ric}(K) \geq \kappa$.
- Finite volume discretisations of Fokker-Planck equations in 1D.

Theorem (Erbar, M. 2012)

Let $(\mathcal{X}_i, K_i, \pi_i)$ be reversible finite Markov chains and let $(\mathcal{X}, K, \pi)$ be the product chain. Then:

$$\mathrm{Ric}(\mathcal{X}_i, K_i, \pi_i) \geq \kappa_i \qquad \Longrightarrow \qquad \mathrm{Ric}(\mathcal{X}, K, \pi) \geq \frac{1}{n} \min_i \kappa_i$$

- Dimension-independent bounds
- Sharp bounds for the discrete hypercube $\{-1, 1\}^n$

Gromov-Hausdorff convergence

- Let $\mathbf{T}_N^d = (\mathbf{Z}/N\mathbf{Z})^d$ be the discrete torus.
- Let $\mathcal{W}_N$ be the normalised transportation metric for simple random walk on $\mathbf{T}_N^d$.

Gromov-Hausdorff convergence

- Let $\mathbf{T}_N^d = (\mathbf{Z}/N\mathbf{Z})^d$ be the discrete torus.
- Let $\mathcal{W}_N$ be the normalised transportation metric for simple random walk on $\mathbf{T}_N^d$.

Theorem (Gigli, M. 2012)

$(\mathcal{P}(\mathbf{T}_N^d), \mathcal{W}_N) \to (\mathcal{P}(\mathbf{T}^d), W_2)$ in the sense of Gromov–Hausdorff.

Gromov-Hausdorff convergence

- Let $\mathbf{T}_N^d = (\mathbf{Z}/N\mathbf{Z})^d$ be the discrete torus.
- Let $\mathcal{W}_N$ be the normalised transportation metric for simple random walk on $\mathbf{T}_N^d$.

Theorem (Gigli, M. 2012)

$(\mathcal{P}(\mathbf{T}_N^d), \mathcal{W}_N) \to (\mathcal{P}(\mathbf{T}^d), W_2)$ in the sense of Gromov–Hausdorff.

- Compatibility between W_2 and $\mathcal{W}$.

Gromov-Hausdorff convergence

- Let $\mathbf{T}_N^d = (\mathbf{Z}/N\mathbf{Z})^d$ be the discrete torus.
- Let $\mathcal{W}_N$ be the normalised transportation metric for simple random walk on $\mathbf{T}_N^d$.

Theorem (Gigli, M. 2012)

$(\mathcal{P}(\mathbf{T}_N^d), \mathcal{W}_N) \to (\mathcal{P}(\mathbf{T}^d), W_2)$ in the sense of Gromov–Hausdorff.

- Compatibility between W_2 and $\mathcal{W}$.
- Main ingredient for proving convergence of gradient flows.

Further developments

Closely related gradient flow structures have been discovered for

Further developments

Closely related gradient flow structures have been discovered for

- Systems of chemical reactions (Mielke)
  non-linear generalisation of continuous time Markov chains

Further developments

Closely related gradient flow structures have been discovered for

- Systems of chemical reactions (Mielke)
  non-linear generalisation of continuous time Markov chains
- Non-local equations on general state spaces (Erbar)
  fractional heat equations

Further developments

Closely related gradient flow structures have been discovered for

- Systems of chemical reactions (Mielke)
  non-linear generalisation of continuous time Markov chains
- Non-local equations on general state spaces (Erbar)
  fractional heat equations
- Discrete porous medium equations (Erbar-M.)
  allows for structure-preserving discretisations of PDEs

Further developments

Closely related gradient flow structures have been discovered for

- Systems of chemical reactions (Mielke)
  non-linear generalisation of continuous time Markov chains
- Non-local equations on general state spaces (Erbar)
  fractional heat equations
- Discrete porous medium equations (Erbar-M.)
  allows for structure-preserving discretisations of PDEs
- Dissipative quantum mechanics (Carlen-M., Mielke)
  non-commutative analogue of $\mathcal{W}$ for density matrices

Thank you!