Ricci curvature of Markov chains via convexity of the entropy

Jan Maas
University of Bonn

6ICSA Bedlewo
10 September 2012

Joint work with Matthias Erbar (Bonn)

Starting point: Displacement convexity of the entropy

Connection between:

- Boltzmann-Shannon entropy:

$$\mathrm{Ent}(\mu) = \int_{\mathbf{R}^n} \rho(x) \log \rho(x) \, \mathsf{d}x; \qquad \frac{d\mu}{dx} = \rho$$

Starting point: Displacement convexity of the entropy

Connection between:

- Boltzmann-Shannon entropy:

$$\mathrm{Ent}(\mu) = \int_{\mathbf{R}^n} \rho(x) \log \rho(x) \, \mathsf{d}x; \qquad \frac{d\mu}{dx} = \rho$$

- L^2-Wasserstein metric:

$$W_2(\mu_0, \mu_1)^2 = \inf \left\{ \int_{\mathbf{R}^n \times \mathbf{R}^n} |x - y|^2 \, \mathsf{d}\gamma(x, y) \right. \\ \left. : \gamma \text{ with marginals } \mu_0 \text{ and } \mu_1 \right\}$$

Starting point: Displacement convexity of the entropy

Connection between:

- Boltzmann-Shannon entropy:

$$\mathrm{Ent}(\mu) = \int_{\mathbf{R}^n} \rho(x) \log \rho(x)\, \mathsf{d}x; \qquad \frac{d\mu}{dx} = \rho$$

- L^2-Wasserstein metric:

$$W_2(\mu_0, \mu_1)^2 = \inf \left\{ \int_{\mathbf{R}^n \times \mathbf{R}^n} |x - y|^2\, \mathsf{d}\gamma(x, y) \right. \\ \left. : \gamma \text{ with marginals } \mu_0 \text{ and } \mu_1 \right\}$$

Theorem (McCann '94)

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

Starting Point: Ricci curvature and optimal transport

Theorem (Otto, Villani; Cordero-Erausquin, McCann, Schmuckenschläger; von Renesse, Sturm)

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

1. $\mathrm{Ric} \geq \kappa$ everywhere on $\mathcal{M}$
2. Displacement κ-convexity of the entropy, i.e.,

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

   for all L^2-Wasserstein geodesics $(\mu_t)_{t\in[0,1]}$ in $\mathcal{P}(\mathcal{M})$.

Synthetic Ricci curvature of metric measure spaces

Definition (Lott, Villani; Sturm)

A metric measure space $(\mathcal{X}, d, m)$ satisfies $CD(\kappa, \infty)$ if any $\mu_0, \mu_1 \in \mathcal{P}_2(\mathcal{X})$ can be connected by a constant speed W_2-geodesic $(\mu_t)_{t\in[0,1]}$ such that

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

Synthetic Ricci curvature of metric measure spaces

Definition (Lott, Villani; Sturm)

A metric measure space $(\mathcal{X}, d, m)$ satisfies $CD(\kappa, \infty)$ if any $\mu_0, \mu_1 \in \mathcal{P}_2(\mathcal{X})$ can be connected by a constant speed W_2-geodesic $(\mu_t)_{t\in[0,1]}$ such that

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

Crucial features:

- Many geometric, analytic and probabilistic consequences
  - $\longrightarrow$ (log-)Sobolev inequalities, Talagrand inequalities, Brunn–Minkowski, etc.

Synthetic Ricci curvature of metric measure spaces

Definition (Lott, Villani; Sturm)

A metric measure space $(\mathcal{X}, d, m)$ satisfies $CD(\kappa, \infty)$ if any $\mu_0, \mu_1 \in \mathcal{P}_2(\mathcal{X})$ can be connected by a constant speed W_2-geodesic $(\mu_t)_{t\in[0,1]}$ such that

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

Crucial features:

- Many geometric, analytic and probabilistic consequences
  $\longrightarrow$ (log-)Sobolev inequalities, Talagrand inequalities, Brunn–Minkowski, etc.
- Stability under measured Gromov–Hausdorff convergence

Synthetic Ricci curvature of metric measure spaces

Definition (Lott, Villani; Sturm)

A metric measure space $(\mathcal{X}, d, m)$ satisfies $CD(\kappa, \infty)$ if any $\mu_0, \mu_1 \in \mathcal{P}_2(\mathcal{X})$ can be connected by a constant speed W_2-geodesic $(\mu_t)_{t\in[0,1]}$ such that

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

Crucial features:

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

Synthetic Ricci curvature of metric measure spaces

Definition (Lott, Villani; Sturm)

A metric measure space $(\mathcal{X}, d, m)$ satisfies $CD(\kappa, \infty)$ if any $\mu_0, \mu_1 \in \mathcal{P}_2(\mathcal{X})$ can be connected by a constant speed W_2-geodesic $(\mu_t)_{t\in[0,1]}$ such that

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

Crucial features:

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

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 $\alpha \in [0, 1]$, and note that

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

What about discrete spaces?

- Example: 2-point space $\mathcal{X} = \{0, 1\}$.
- Set $\mu_\alpha := (1 - \alpha)\delta_0 + \alpha\delta_1$ for $\alpha \in [0, 1]$, and note that

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

- 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 $\alpha \in [0, 1]$, and note that

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

- 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.

- Conclusion: there are no non-trivial W_2-geodesics.

What about discrete spaces?

- Example: 2-point space $\mathcal{X} = \{0, 1\}$.
- Set $\mu_\alpha := (1-\alpha)\delta_0 + \alpha\delta_1$ for $\alpha \in [0,1]$, and note that

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

- 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.

- Conclusion: 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 $\alpha \in [0,1]$, and note that

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

- 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.

- Conclusion: 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.

LSV-definition does not apply to discrete spaces.

Can we adapt LSV to the discrete case?

Why Wasserstein?

Can we adapt LSV to the discrete case?

Why Wasserstein?

Theorem (Jordan, Kinderlehrer, Otto '98)

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

Can we adapt LSV to the discrete case?

Why 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?

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

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

Can we adapt LSV to the discrete case?

Why Wasserstein?

Theorem (Jordan, Kinderlehrer, Otto '98)

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

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

How to make sense of gradient flows in metric spaces?

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

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

Heat flow is gradient flows 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 is gradient flows 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, i.e., $\forall x \;:\; \sum_y K(x,y) = 1$

Discrete setting

Setting

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

Assumptions

- K is irreducible $\longrightarrow$ $\exists!$ inv. measure π
- π is reversible i.e., $\forall x, y \; : \; K(x,y)\pi(x) = K(y,x)\pi(y)$

Discrete setting

Setting

- $\mathcal{X}$: finite set
- $K : \mathcal{X} \times \mathcal{X} \to \mathbf{R}_+$ Markov kernel, i.e., $\forall x \ : \ \sum_y K(x,y) = 1$

Assumptions

- K is irreducible $\longrightarrow$ $\exists!$ inv. measure π
- π is reversible i.e., $\forall x, y \ : \ K(x,y)\pi(x) = K(y,x)\pi(y)$

Heat flow

- $H(t) = e^{t(K-I)}$ is the continuous time Markov semigroup

Discrete setting

Setting

- $\mathcal{X}$: finite set
- $K : \mathcal{X} \times \mathcal{X} \to \mathbf{R}_+$ Markov kernel, i.e., $\forall x \ : \ \sum_y K(x,y) = 1$

Assumptions

- K is irreducible $\longrightarrow$ $\exists!$ inv. measure π
- π is reversible i.e., $\forall x, y \ : \ K(x,y)\pi(x) = K(y,x)\pi(y)$

Heat flow

- $H(t) = e^{t(K-I)}$ is the continuous time Markov semigroup

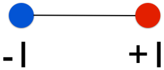
Relative Entropy

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

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

Simplest non-trivial example: 2-point space

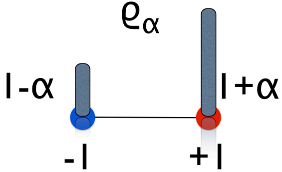
$\mathcal{X} = \{-1, 1\}$
$K(-1, 1) = K(1, -1) = 1$
$\pi(-1) = \pi(1) = \frac{1}{2}$



Simplest non-trivial example: 2-point space

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

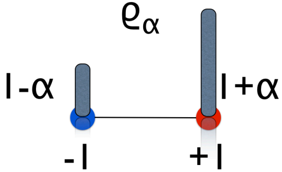
$\rightsquigarrow$ Every $\rho \in \mathcal{P}(\mathcal{X})$ is of the form ρ_α, $\alpha \in [-1, 1]$.



Simplest non-trivial example: 2-point space

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

$\rightsquigarrow$ Every $\rho \in \mathcal{P}(\mathcal{X})$ is of the form ρ_α, $\alpha \in [-1, 1]$.



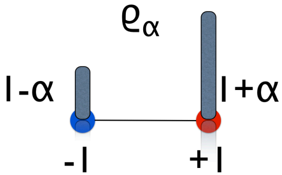
Question

Is the heat flow the gradient flow of Ent w.r.t the L^2-Wasserstein metric?

Simplest non-trivial example: 2-point space

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

$\rightsquigarrow$ Every $\rho \in \mathcal{P}(\mathcal{X})$ is of the form ρ_α, $\alpha \in [-1, 1]$.



Question

Is the heat flow the gradient flow of Ent w.r.t the L^2-Wasserstein metric?

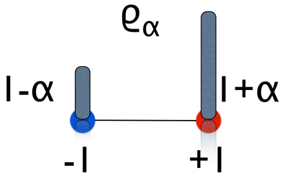
Answer

NO! Reason:
$W_2(\rho_\alpha, \rho_\beta) = \sqrt{2|\beta - \alpha|}$.

Simplest non-trivial example: 2-point space

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

$\rightsquigarrow$ Every $\rho \in \mathcal{P}(\mathcal{X})$ is of the form ρ_α, $\alpha \in [-1, 1]$.



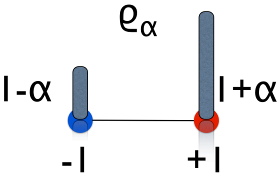
Question

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

Simplest non-trivial example: 2-point space

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

$\rightsquigarrow$ Every $\rho \in \mathcal{P}(\mathcal{X})$ is of the form ρ_α, $\alpha \in [-1, 1]$.



Question

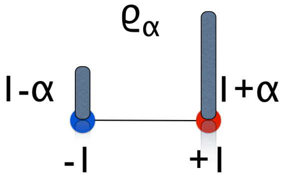
Is the heat flow the gradient flow of Ent w.r.t some other metric on $\mathcal{P}(\{-1,1\})$?

Answer

YES!

Simplest non-trivial example: 2-point space

$\mathcal{X} = \{-1, 1\}$
$K(-1, 1) = K(1, -1) = 1$
$\pi(-1) = \pi(1) = \frac{1}{2}$
$\rightsquigarrow$ Every $\rho \in \mathcal{P}(\mathcal{X})$ is of the form ρ_α, $\alpha \in [-1, 1]$.



Question

Is the heat flow the gradient flow of Ent 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

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

How to define $\mathcal{W}$ in the general discrete case?

In $\mathbf{R}^n$ there is a dynamical characterisation of W_2:

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

$$W_2(\bar{\rho}_0, \bar{\rho}_1)^2 = \inf_{\rho_\cdot, \Psi_\cdot} \left\{ \int_0^1 \int_{\mathbf{R}^n} |\Psi_t(x)|^2 \, \rho_t(x) \, \mathrm{d}x \, \mathrm{d}t \;\; : \right.$$
$$\left. \partial_t \rho + \nabla \cdot (\rho \Psi) = 0 \; , \quad \rho_0 = \bar{\rho}_0 \; , \quad \rho_1 = \bar{\rho}_1 \right\} .$$

How to define $\mathcal{W}$ in the general discrete case?

In $\mathbf{R}^n$ there is a dynamical characterisation of W_2:

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

$$W_2(\bar\rho_0,\bar\rho_1)^2 = \inf_{\rho_\cdot,\Psi_\cdot} \left\{ \int_0^1 \int_{\mathbf{R}^n} |\Psi_t(x)|^2 \rho_t(x)\,\mathrm{d}x\,\mathrm{d}t \; : \; \partial_t\rho + \nabla\cdot(\rho\Psi) = 0\ , \quad \rho_0 = \bar\rho_0\ , \quad \rho_1 = \bar\rho_1 \right\} .$$

- One may restrict to gradients: $\Psi_t = \nabla\psi_t$

How to define $\mathcal{W}$ in the general discrete case?

In $\mathbf{R}^n$ there is a dynamical characterisation of W_2:

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

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

- One may restrict to gradients: $\Psi_t = \nabla \psi_t$

How to define $\mathcal{W}$ in the general discrete case?

In $\mathbf{R}^n$ there is a dynamical characterisation of W_2:

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

$$W_2(\bar\rho_0, \bar\rho_1)^2 = \inf_{\rho_\cdot, \psi_\cdot} \bigg\{ \int_0^1 \int_{\mathbf{R}^n} |\nabla \psi_t(x)|^2 \rho_t(x) \, \mathrm{d}x \, \mathrm{d}t \ : \ \partial_t \rho + \nabla \cdot (\rho \nabla \psi) = 0 \ , \quad \rho_0 = \bar\rho_0 \ , \quad \rho_1 = \bar\rho_1 \bigg\} \ .$$

- One may restrict to gradients: $\Psi_t = \nabla \psi_t$
- Otto's interpretation: Riemannian distance formula

How to define $\mathcal{W}$ in the general discrete case?

In $\mathbf{R}^n$ there is a dynamical characterisation of W_2:

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

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

- One may restrict to gradients: $\Psi_t = \nabla \psi_t$
- Otto's interpretation: Riemannian distance formula
- Idea: *define* a metric in the discrete case using this formula.

How to define $\mathcal{W}$ in the general discrete case?

In $\mathbf{R}^n$ there is a dynamical characterisation of W_2:

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

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

- One may restrict to gradients: $\Psi_t = \nabla\psi_t$
- Otto's interpretation: Riemannian distance formula

- Idea: *define* a metric in the discrete case using this formula.
- Obstruction: how to multiply probability densities and discrete gradients?

Definition of $\mathcal{W}$

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

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

$$\text{s.t.} \qquad \partial_t\rho + \nabla\cdot(\rho\nabla\psi) = 0 \ .$$

Definition of $\mathcal{W}$

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

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

$$\text{s.t.} \qquad \partial_t \rho + \nabla \cdot (\rho \nabla \psi) = 0 \; .$$

Definition in the discrete case (M. 2011)

$$\mathcal{W}(\bar\rho_0, \bar\rho_1)^2 := \inf_{\rho, \psi} \left\{ \qquad \right\}$$

s.t.

$$\rho_0 = \bar\rho_0 \; , \quad \rho_1 = \bar\rho_1 \; .$$

Definition of $\mathcal{W}$

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

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

$$\text{s.t.} \qquad \partial_t \rho + \nabla \cdot (\rho \nabla \psi) = 0 \ .$$

Definition in the discrete case (M. 2011)

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

s.t.

$$\rho_0 = \bar\rho_0 \ , \quad \rho_1 = \bar\rho_1 \ .$$

Definition of $\mathcal{W}$

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

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

$$\text{s.t.} \qquad \partial_t \rho + \nabla \cdot (\rho \nabla \psi) = 0 \; .$$

Definition in the discrete case (M. 2011)

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

s.t.

$$\rho_0 = \bar\rho_0 \; , \quad \rho_1 = \bar\rho_1 \; .$$

Definition of $\mathcal{W}$

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

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

$$\text{s.t.} \qquad \partial_t \rho + \nabla \cdot (\rho \nabla \psi) = 0 \; .$$

Definition in the discrete case (M. 2011)

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

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

s.t.

$$\rho_0 = \bar\rho_0 \; , \quad \rho_1 = \bar\rho_1 \; .$$

Definition of $\mathcal{W}$

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

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

$$\text{s.t.} \qquad \partial_t \rho + \nabla \cdot (\rho \nabla \psi) = 0 \; .$$

Definition in the discrete case (M. 2011)

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

s.t.

$$\rho_0 = \bar\rho_0 \; , \quad \rho_1 = \bar\rho_1 \; .$$

Definition of $\mathcal{W}$

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

$$W_2^2(\bar{\rho}_0, \bar{\rho}_1) = \inf_{\rho_\cdot, \psi_\cdot} \left\{ \int_0^1 \int_{\mathbf{R}^n} |\nabla \psi_t(x)|^2 \, \rho_t(x) \, \mathrm{d}x \, \mathrm{d}t \right\}$$

$$\text{s.t.} \qquad \partial_t \rho + \nabla \cdot (\rho \nabla \psi) = 0 \; .$$

Definition in the discrete case (M. 2011)

$$\mathcal{W}(\bar{\rho}_0, \bar{\rho}_1)^2$$

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

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

$$\rho_0 = \bar{\rho}_0 \; , \quad \rho_1 = \bar{\rho}_1 \; .$$

Definition of $\mathcal{W}$

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

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

$$\text{s.t.} \qquad \partial_t \rho + \nabla \cdot (\rho \nabla \psi) = 0 \ .$$

Definition in the discrete case (M. 2011)

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

$$\text{s.t.} \qquad \frac{\mathrm{d}}{\mathrm{d}t} \rho_t(x) + \sum_{y \in \mathcal{X}} \hat\rho_t(x,y)(\psi_t(x) - \psi_t(y)) K(x,y) = 0 \qquad \forall x,$$
$$\rho_0 = \bar\rho_0 \ , \quad \rho_1 = \bar\rho_1 \ .$$

How should we define $\hat\rho$?

Definition of $\mathcal{W}$

Definition in the discrete case

$$\mathcal{W}(\bar{\rho}_0, \bar{\rho}_1)^2$$

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

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

Definition of $\mathcal{W}$

Definition in the discrete case

$$\mathcal{W}(\bar\rho_0,\bar\rho_1)^2$$
$$:= \inf_{\rho_\cdot,\psi_\cdot} \left\{ \frac{1}{2}\int_0^1 \sum_{x,y\in\mathcal{X}} \big(\psi_t(x)-\psi_t(y)\big)^2 \hat\rho_t(x,y)K(x,y)\pi(x)\,\mathrm{d}t \right\}$$
$$\text{s.t.} \quad \frac{\mathrm{d}}{\mathrm{d}t}\rho_t(x) + \sum_{y\in\mathcal{X}} (\psi_t(x)-\psi_t(y))\hat\rho_t(x,y)K(x,y) = 0\,.$$

How should we define $\hat\rho$? $\longrightarrow$ logarithmic mean

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

Results

Theorem (M. 2011)

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

Results

Theorem (M. 2011)

1. $\mathcal{W}$ defines metric on $\mathcal{P}(\mathcal{X})$.
2. The space $\{\rho \in \mathcal{P}(\mathcal{X}) \; : \; \rho(x) > 0 \; \forall x\}$ is a Riemannian manifold with metric $\mathcal{W}$.

Results

Theorem (M. 2011)

1. $\mathcal{W}$ defines metric on $\mathcal{P}(\mathcal{X})$.
2. The space $\{\rho \in \mathcal{P}(\mathcal{X}) \ : \ \rho(x) > 0 \ \forall x\}$ is a Riemannian manifold with metric $\mathcal{W}$.
3. 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

Theorem (M. 2011)

1. $\mathcal{W}$ defines metric on $\mathcal{P}(\mathcal{X})$.
2. The space $\{\rho \in \mathcal{P}(\mathcal{X}) \;:\; \rho(x) > 0 \;\forall x\}$ is a Riemannian manifold with metric $\mathcal{W}$.
3. The tangent space at ρ is the set of discrete gradients with

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

4. The heat flow is the gradient flow of the entropy.

Results

Theorem (M. 2011)

1. $\mathcal{W}$ defines metric on $\mathcal{P}(\mathcal{X})$.
2. The space $\{\rho \in \mathcal{P}(\mathcal{X}) \ : \ \rho(x) > 0 \ \forall x\}$ is a Riemannian manifold with metric $\mathcal{W}$.
3. The tangent space at ρ is the set of discrete gradients with

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

4. The heat flow is the gradient flow of the entropy.

Remark

Related independent work by

- Chow, Huang, Li, and Zhou
- Mielke

Why the logarithmic mean?

Formal proof of the JKO-Theorem

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

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

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

Why the logarithmic mean?

Formal proof of the JKO-Theorem

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

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

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

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

$$\partial_t \rho = \nabla \cdot (\nabla \rho) = -\nabla \cdot (\rho \nabla \psi) \ .$$

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. $\partial_t \rho + \nabla \cdot (\rho \nabla \psi) = 0$, then

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

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

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

$$\partial_t \rho = \nabla \cdot (\nabla \rho) = -\nabla \cdot (\rho \nabla \psi) \ .$$

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

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

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

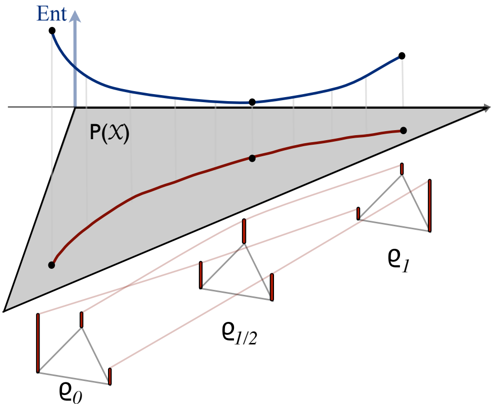
$$\rho(x) - \rho(y) = \hat{\rho}(x,y)\big(\log \rho(x) - \log \rho(y)\big)$$

Ricci curvature of Markov chains

The discrete analogue of Lott–Sturm–Villani becomes:

Definition (Erbar, M. 2011)

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

Theorem (Erbar, M.)

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

1. à la Bakry–Émery: $\mathrm{Ric}(K) \geq \kappa \Longrightarrow$ modified log-Sobolev, i.e.

$$\mathrm{Ent}(\rho) \leq \frac{1}{2\kappa}\mathcal{E}(\rho, \log \rho)\ .$$

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

Consequences: Sharp functional inequalities

Theorem (Erbar, M.)

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

1. à la Bakry–Émery: $\mathrm{Ric}(K) \geq \kappa \Longrightarrow$ modified log-Sobolev, i.e.

$$\mathrm{Ent}(\rho) \leq \frac{1}{2\kappa}\mathcal{E}(\rho, \log \rho) \; .$$

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

2. à la Otto–Villani: mod. log-Sobolev $\Longrightarrow$ mod. Talagrand, i.e.

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

Consequences: Sharp functional inequalities

Theorem (Erbar, M.)

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

1. à la Bakry–Émery: $\mathrm{Ric}(K) \geq \kappa \Longrightarrow$ modified log-Sobolev, i.e.

$$\mathrm{Ent}(\rho) \leq \frac{1}{2\kappa}\mathcal{E}(\rho, \log \rho)\ .$$

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

2. à la Otto–Villani: mod. log-Sobolev $\Longrightarrow$ mod. Talagrand, i.e.

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

3. mod. Talagrand $\Longrightarrow$ [spectral gap and T_1]:

$$\|\varphi\|^2_{L^2(\mathcal{X},\pi)} \leq \frac{1}{\kappa}\mathcal{E}(\varphi, \varphi) \quad \text{and} \quad W_1(\rho, \mathbf{1})^2 \leq \frac{1}{\kappa}\,\mathrm{Ent}(\rho)\ .$$

Ricci bounds: examples

- (Mielke 2012) For every finite reversible Markov chain: $\exists \kappa \in \mathbf{R}$ such that $\mathrm{Ric}(K) \geq \kappa$.

Ricci bounds: examples

- (Mielke 2012) For every finite reversible Markov chain: $\exists \kappa \in \mathbf{R}$ such that $\mathrm{Ric}(K) \geq \kappa$.
- (Mielke 2012) Finite volume discretisations of Fokker-Planck equations in 1D

Ricci bounds: examples

- (Mielke 2012) For every finite reversible Markov chain: $\exists \kappa \in \mathbf{R}$ such that $\mathrm{Ric}(K) \geq \kappa$.
- (Mielke 2012) 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

- (Mielke 2012) For every finite reversible Markov chain: $\exists \kappa \in \mathbf{R}$ such that $\mathrm{Ric}(K) \geq \kappa$.
- (Mielke 2012) 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

- (Mielke 2012) For every finite reversible Markov chain: $\exists \kappa \in \mathbf{R}$ such that $\mathrm{Ric}(K) \geq \kappa$.
- (Mielke 2012) 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 renormalised 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 renormalised 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 renormalised 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

- Systems of reaction-diffusion equations (Mielke)
- Fractional heat equations (Erbar)
- Dissipative quantum mechanics (Carlen, M.; Mielke)

Further developments

- Systems of reaction-diffusion equations (Mielke)
- Fractional heat equations (Erbar)
- Dissipative quantum mechanics (Carlen, M.; Mielke)

Thank you!