Structure Preserving Discretizations of Wasserstein Gradient Flows
Part I of IV: The object to discretize

Daniel Matthes (TU Munich)

"A school on gradient flows"
ICTS Bengaluru 7.–11.9.2026

Where do we need to go today?

Question

What is that alleged interesting structure of a Wasserstein gradient flow that a discretization should preserve?

Let us rewrite Wasserstein as Riemann.

An ODE analogy for gradient flows (1/3)

A function $\boldsymbol{e} : \mathbb{R}^n \to \mathbb{R}$ defines an euclidean gradient flow via

$$\dot{x} = -\nabla \boldsymbol{e}(x).$$

An ODE analogy for gradient flows (1/3)

A function $\boldsymbol{e} : \mathbb{R}^n \to \mathbb{R}$ defines an euclidean gradient flow via

$$\dot{x} = -\nabla \boldsymbol{e}(x).$$

Solutions $(x_t)_{t \geq 0}$ diminish $\boldsymbol{e}$ "as fast as possible":

$$-\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{e}(x_t) = -\dot{x}_t \cdot \nabla \boldsymbol{e}(x_t) = \|\dot{x}_t\|^2 = \|\nabla \boldsymbol{e}(x_t)\|^2,$$

i.e., $(x_t)_{t \geq 0}$ invests all of its velocity into dissipating $\boldsymbol{e}$.

An ODE analogy for gradient flows (2/3)

On a manifold $\mathcal{M}$ the tangent space $\mathrm{T}_x\mathcal{M}$ at $x \in \mathcal{M}$ is formed by velocities $v \cong \dot{x}_0$ of curves $(x_s)_{s\in(-\sigma,\sigma)}$ with $x_0 = x$.

An ODE analogy for gradient flows (2/3)

On a manifold $\mathcal{M}$ the tangent space $\mathrm{T}_x\mathcal{M}$ at $x \in \mathcal{M}$ is formed by velocities $v \cong \dot{x}_0$ of curves $(x_s)_{s\in(-\sigma,\sigma)}$ with $x_0 = x$.

A Riemannian tensor assigns to each $x \in \mathcal{M}$ a scalar product $\langle\cdot,\cdot\rangle_x$. It defines a norm on $\mathrm{T}_x\mathcal{M}$ and a distance on $\mathcal{M}$:

$$\|v\|_x^2 = \langle v, v\rangle_x \quad \text{and} \quad \mathbf{d}(x_0, x_1)^2 = \inf_{(x_s)_{s\in[0,1]}} \int_0^1 \|v_s\|_{x_s}^2 \,\mathrm{d}s.$$

An ODE analogy for gradient flows (2/3)

On a manifold $\mathcal{M}$ the tangent space $\mathrm{T}_x\mathcal{M}$ at $x \in \mathcal{M}$ is formed by velocities $v \cong \dot{x}_0$ of curves $(x_s)_{s\in(-\sigma,\sigma)}$ with $x_0 = x$.

A Riemannian tensor assigns to each $x \in \mathcal{M}$ a scalar product $\langle \cdot, \cdot \rangle_x$.
It defines a norm on $\mathrm{T}_x\mathcal{M}$ and a distance on $\mathcal{M}$:

$$\|v\|_x^2 = \langle v, v \rangle_x \quad \text{and} \quad \mathbf{d}(x_0, x_1)^2 = \inf_{(x_s)_{s\in[0,1]}} \int_0^1 \|v_s\|_{x_s}^2 \, \mathrm{d}s.$$

The differential $\mathrm{D}_x \boldsymbol{e}$ of a function $\boldsymbol{e} : \mathcal{M} \to \mathbb{R}$ assigns to each $v \in \mathrm{T}_x\mathcal{M}$ the directional derivative,

$$\mathrm{D}_x \boldsymbol{e}[v] = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \boldsymbol{e}(x_s) \quad \text{with} \quad v \cong \dot{x}_0.$$

An ODE analogy for gradient flows (2/3)

On a manifold $\mathcal{M}$ the tangent space $\mathrm{T}_x\mathcal{M}$ at $x \in \mathcal{M}$ is formed by velocities $v \cong \dot{x}_0$ of curves $(x_s)_{s\in(-\sigma,\sigma)}$ with $x_0 = x$.

A Riemannian tensor assigns to each $x \in \mathcal{M}$ a scalar product $\langle\cdot,\cdot\rangle_x$. It defines a norm on $\mathrm{T}_x\mathcal{M}$ and a distance on $\mathcal{M}$:

$$\|v\|_x^2 = \langle v, v\rangle_x \quad \text{and} \quad \mathbf{d}(x_0, x_1)^2 = \inf_{(x_s)_{s\in[0,1]}} \int_0^1 \|v_s\|_{x_s}^2 \,\mathrm{d}s.$$

The differential $\mathrm{D}_x\boldsymbol{e}$ of a function $\boldsymbol{e} : \mathcal{M} \to \mathbb{R}$ assigns to each $v \in \mathrm{T}_x\mathcal{M}$ the directional derivative,

$$\mathrm{D}_x\boldsymbol{e}[v] = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \boldsymbol{e}(x_s) \quad \text{with} \quad v \cong \dot{x}_0.$$

The gradient $\nabla_x\boldsymbol{e}$ is the Riesz representative of $\mathrm{D}_x\boldsymbol{e}$ w.r.t. $\langle\cdot,\cdot\rangle_x$,

$$\langle\nabla_x\boldsymbol{e}, v\rangle_x = \mathrm{D}_x\boldsymbol{e}[v] \quad \text{for all } v \in \mathrm{T}_x\mathcal{M}.$$

An ODE analogy for gradient flows (3/3)

Accordingly, $\boldsymbol{e} : \mathcal{M} \to \mathbb{R}$ defines a Riemannian gradient flow via

$$\dot{x} \cong -\nabla_x \boldsymbol{e}.$$

By definition, solutions $(x_t)_{t \geq 0}$ satisfy

$$-\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{e}(x_t) = -\mathrm{D}_{x_t}\boldsymbol{e}[v_t] = \langle -\nabla_{x_t}\boldsymbol{e}, v_t \rangle_{x_t} = \left\| \nabla_{x_t}\boldsymbol{e} \right\|_{x_t}^2 = \left\| v_t \right\|_{x_t}^2,$$

i.e., all velocity is invested into dissipating $\boldsymbol{e}$.

Otto calculus (1/3)

On the "manifold" of probability densities,

$$\mathcal{P}_2(\mathbb{R}^d) = \left\{ \rho \in L^1(\mathbb{R}^d) \,\middle|\, \rho \geq 0,\ \int \rho \, \mathrm{d}x = 1, \int |x|^2 \rho \, \mathrm{d}x < \infty \right\},$$

the L^2-Wasserstein metric is a distance

$$\mathbb{W}(\rho_0, \rho_1)^2 = \int \left\{ \int |T(x) - x|^2 \rho_0(x) \, \mathrm{d}x \,\middle|\, T\#\rho_0 = \rho_1 \right\}.$$

Can one define a Riemannian metric on $\mathcal{M} = \mathcal{P}_2(\mathbb{R}^d)$ such that $\mathbf{d} = \mathbb{W}$ (…and make the according gradient flow explicit)?

Otto calculus (2/3)

Otto's tangent space

The tangent space $\mathrm{T}_\rho \mathcal{P}_2(\mathbb{R}^d)$ is identified with functions $\varphi : \mathbb{R}^d \to \mathbb{R}$ (modulo constants) such that

$$\|\varphi\|_\rho^2 := \int |\nabla \varphi|^2 \rho \, \mathrm{d}x < \infty.$$

The tangent vector $\varphi \in \mathrm{T}_\rho \mathcal{P}_2(\mathbb{R}^d)$ relates to curves $(\rho_s)_{s \in (-\sigma, \sigma)}$ s.t.

$$\dot\rho_s = -\operatorname{div}(\rho_s \nabla \varphi_s) \quad \text{with } \rho_0 = \rho,\ \varphi_0 = \varphi.$$

Otto calculus (2/3)

Otto's tangent space

The tangent space $\mathrm{T}_\rho \mathcal{P}_2(\mathbb{R}^d)$ is identified with functions $\varphi : \mathbb{R}^d \to \mathbb{R}$ (modulo constants) such that

$$\|\varphi\|_\rho^2 := \int |\nabla \varphi|^2 \rho \, \mathrm{d}x < \infty.$$

The tangent vector $\varphi \in \mathrm{T}_\rho \mathcal{P}_2(\mathbb{R}^d)$ relates to curves $(\rho_s)_{s \in (-\sigma, \sigma)}$ s.t.

$$\dot{\rho}_s = -\operatorname{div}(\rho_s \nabla \varphi_s) \quad \text{with } \rho_0 = \rho,\ \varphi_0 = \varphi.$$

Theorem ([Benamou&Brenier'00])

$$\mathbb{W}(\rho_0, \rho_1)^2 = \inf_{\rho_s} \int_0^1 \|\varphi_s\|_{\rho_s}^2 \, \mathrm{d}s.$$

Otto calculus (3/3)

The norm $\|\cdot\|_\rho$ induces the metric tensor

$$\langle \varphi, \psi \rangle_\rho = \int \nabla\varphi \cdot \nabla\psi \, \rho \, \mathrm{d}x.$$

The gradient ψ of $\boldsymbol{E}$ at ρ is thus characterized by

$$\int \nabla\varphi \cdot \nabla\psi \, \rho \, \mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} \boldsymbol{E}(\rho_s) \quad \text{with} \quad \dot{\rho}_s = -\operatorname{div}(\rho\nabla\varphi).$$

Otto calculus (3/3)

The norm $\|\cdot\|_\rho$ induces the metric tensor

$$\langle \varphi, \psi \rangle_\rho = \int \nabla\varphi \cdot \nabla\psi \, \rho \, \mathrm{d}x.$$

The gradient ψ of $\boldsymbol{E}$ at ρ is thus characterized by

$$\int \nabla\varphi \cdot \nabla\psi \, \rho \, \mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} \boldsymbol{E}(\rho_s) \quad \text{with} \quad \dot{\rho}_s = -\operatorname{div}(\rho\nabla\varphi).$$

The differential of a functional $\boldsymbol{E} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ at ρ is is implicitly defined via $\delta\boldsymbol{E}/\delta\rho : \mathbb{R}^d \to \mathbb{R}$ such that

$$\int \zeta \, \frac{\delta \boldsymbol{E}}{\delta\rho} \, \mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} \boldsymbol{E}(\rho + s\zeta) \quad \text{for all suitable } \zeta : \mathbb{R}^d \to \mathbb{R}.$$

Otto calculus (3/3)

The norm $\|\cdot\|_\rho$ induces the metric tensor

$$\langle \varphi, \psi \rangle_\rho = \int \nabla\varphi \cdot \nabla\psi \, \rho \, \mathrm{d}x.$$

The gradient ψ of $\boldsymbol{E}$ at ρ is thus characterized by

$$\int \nabla\varphi \cdot \nabla\psi \, \rho \, \mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} \boldsymbol{E}(\rho_s) \quad \text{with} \quad \dot{\rho}_s = -\operatorname{div}(\rho \nabla \varphi).$$

The differential of a functional $\boldsymbol{E} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ at ρ is
is implicitly defined via $\delta \boldsymbol{E}/\delta\rho : \mathbb{R}^d \to \mathbb{R}$ such that

$$\int \zeta \, \frac{\delta \boldsymbol{E}}{\delta \rho} \, \mathrm{d}x = \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} \boldsymbol{E}(\rho + s\zeta) \quad \text{for all suitable } \zeta : \mathbb{R}^d \to \mathbb{R}.$$

Gradient flow equation

$$\partial_t \rho = \operatorname{div}\left(\rho \, \nabla \left(\frac{\delta \boldsymbol{E}}{\delta \rho}\right)\right).$$

Does it get even more abstract?

There are interesting examples.

Standard example 1: Continuity equation

Potential energy

For an external potential $V : \mathbb{R}^d \to \mathbb{R}$ consider

$$\boldsymbol{E}(\rho) = \int V(x)\rho(x)\,\mathrm{d}x.$$

Then $\delta \boldsymbol{E}/\delta\rho = V$, and accordingly,
the $\mathbb{W}$-gradient flow is the continuity equation,

$$\partial_t \rho = \operatorname{div}\big(\rho\,\nabla V\big).$$

Standard example 1: Continuity equation

Potential energy

For an external potential $V : \mathbb{R}^d \to \mathbb{R}$ consider

$$\boldsymbol{E}(\rho) = \int V(x)\rho(x)\,\mathrm{d}x.$$

Then $\delta \boldsymbol{E}/\delta\rho = V$, and accordingly,
the $\mathbb{W}$-gradient flow is the continuity equation,

$$\partial_t \rho = \operatorname{div}\big(\rho\,\nabla V\big).$$

Let $(x_t)_{t\geq 0}$ satisfy

$$\dot{x}_t = -\nabla V(x_t).$$

Then a distributional solution to the continuity equation is

$$\rho_t = \delta_{x_t}.$$

Standard example 2: Nonlocal aggregation equation

Interaction energy

For an even interaction potential $W : \mathbb{R}^d \to \mathbb{R}$ consider

$$\boldsymbol{E}(\rho) = \frac{1}{2}\int \rho\, W * \rho \,\mathrm{d}x = \frac{1}{2}\iint \rho(x) W(x-x')\rho(x')\,\mathrm{d}x\,\mathrm{d}x'.$$

Then $\delta\boldsymbol{E}/\delta\rho = W * \rho$, and accordingly,
the $\mathbb{W}$-gradient flow is the nonlocal aggregation equation,

$$\partial_t \rho = \mathrm{div}\,\big(\rho\,\nabla W * \rho\big).$$

Standard example 2: Nonlocal aggregation equation

Interaction energy

For an even interaction potential $W : \mathbb{R}^d \to \mathbb{R}$ consider

$$\boldsymbol{E}(\rho) = \frac{1}{2} \int \rho\, W * \rho \, \mathrm{d}x = \frac{1}{2} \iint \rho(x) W(x - x') \rho(x') \, \mathrm{d}x \, \mathrm{d}x'.$$

Then $\delta \boldsymbol{E} / \delta \rho = W * \rho$, and accordingly,
the $\mathbb{W}$-gradient flow is the nonlocal aggregation equation,

$$\partial_t \rho = \operatorname{div} \big(\rho \, \nabla W * \rho \big).$$

Let N particle positions $x_1, \ldots, x_N$ satisfy

$$\dot{x}_k = - \sum_{j \neq k} \nabla W(x_k - x_j).$$

Then a distributional solution to the nonlocal aggregation equation is

$$\rho_t = \frac{1}{N} \sum_k \delta_{x_k(t)}.$$

Standard example 3: Linear diffusion

Boltzmann entropy

Consider the Boltzmann entropy $\boldsymbol{E} := \boldsymbol{H}$, i.e.,

$$\boldsymbol{H}(\rho) = \int \rho \log \rho \, \mathrm{d}x.$$

Then $\delta \boldsymbol{E}/\delta\rho = \log \rho$, and accordingly,
the $\mathbb{W}$-gradient flow is the linear diffusion equation,

$$\partial_t \rho = \mathrm{div}\,\big(\rho \, \nabla \log \rho\big) = \Delta \rho.$$

Standard example 3: Linear diffusion

Boltzmann entropy

Consider the Boltzmann entropy $\boldsymbol{E} := \boldsymbol{H}$, i.e.,

$$\boldsymbol{H}(\rho) = \int \rho \log \rho \,\mathrm{d}x.$$

Then $\delta \boldsymbol{E}/\delta\rho = \log\rho$, and accordingly,
the $\mathbb{W}$-gradient flow is the linear diffusion equation,

$$\partial_t \rho = \operatorname{div}\big(\rho \nabla \log \rho\big) = \Delta\rho.$$

The solution for initial datum $\bar{\rho}$ is given by

$$\rho_t = H_t * \bar{\rho} \quad \text{with} \quad H_t(z) = (4\pi t)^{-d/2} \exp\left(-\frac{|z|^2}{4t}\right).$$

Standard example 4: Nonlinear Diffusion

Internal energy

For a convex internal energy density $h : \mathbb{R}_{\geq 0} \to \mathbb{R}$, consider

$$\boldsymbol{E}(\rho) = \int h(\rho)\,\mathrm{d}x.$$

Then $\delta\boldsymbol{E}/\delta\rho = h'(\rho)$, and accordingly,
the $\mathbb{W}$-gradient flow is the nonlinear diffusion equation,

$$\partial_t \rho = \Delta P(\rho) \quad \text{with} \quad P(r) = rh'(r) - h(r).$$

Standard example 4: Nonlinear Diffusion

Internal energy

For a convex internal energy density $h : \mathbb{R}_{\geq 0} \to \mathbb{R}$, consider

$$\boldsymbol{E}(\rho) = \int h(\rho)\,\mathrm{d}x.$$

Then $\delta \boldsymbol{E}/\delta\rho = h'(\rho)$, and accordingly,
the $\mathbb{W}$-gradient flow is the nonlinear diffusion equation,

$$\partial_t \rho = \Delta P(\rho) \quad \text{with} \quad P(r) = rh'(r) - h(r).$$

Of particular interest: power-type energy resp. homogeneous diffusion,

$$\boldsymbol{E}(\rho) = \int \frac{\rho^m}{m-1}\,\mathrm{d}x \quad \rightsquigarrow \quad \partial_t \rho = \Delta(\rho^m).$$

Standard example 4: Nonlinear Diffusion

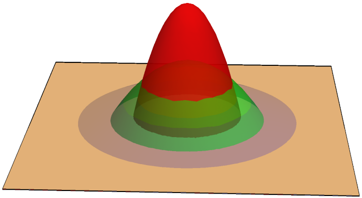
Internal energy

For a convex internal energy density $h : \mathbb{R}_{\geq 0} \to \mathbb{R}$, consider

$$\boldsymbol{E}(\rho) = \int h(\rho)\,\mathrm{d}x.$$

Then $\delta \boldsymbol{E}/\delta\rho = h'(\rho)$, and accordingly,
the $\mathbb{W}$-gradient flow is the nonlinear diffusion equation,

$$\partial_t \rho = \Delta P(\rho) \quad \text{with} \quad P(r) = rh'(r) - h(r).$$



Fourth order example 1: lubrication

Dirichlet energy

Consider the Dirichlet energy

$$\boldsymbol{E}(\rho) = \frac{1}{2}\int |\nabla\rho|^2 \, \mathrm{d}x.$$

Then $\delta\boldsymbol{E}/\delta\rho = -\Delta\rho$, and accordingly,
the $\mathbb{W}$-gradient flow is the thin film equation (a.k.a. Hele-Shaw flow),

$$\partial_t\rho = -\operatorname{div}\big(\rho\,\nabla\Delta\rho\big).$$

Fourth order example 2: quantum diffusion

Fisher information

Consider the Fisher information

$$\boldsymbol{F}(\rho) = 2 \int \left|\nabla \sqrt{\rho}\right|^2 \mathrm{d}x \qquad .$$

Then $\delta \boldsymbol{E}/\delta\rho = -\Delta\sqrt{\rho}/\sqrt{\rho}$, and accordingly,
the $\mathbb{W}$-gradient flow is the DLSS equation (a.k.a. QDD4),

$$\partial_t \rho = -2 \operatorname{div} \left(\rho \, \nabla \left[\frac{\Delta\sqrt{\rho}}{\sqrt{\rho}} \right] \right) \qquad .$$

Fourth order example 2: quantum diffusion

Fisher information

Consider the Fisher information

$$\boldsymbol{F}(\rho) = 2\int \left|\nabla\sqrt{\rho}\right|^2 \mathrm{d}x = \frac{1}{2}\int |\nabla \log\rho|^2 \rho \,\mathrm{d}x.$$

Then $\delta \boldsymbol{E}/\delta\rho = -\Delta\sqrt{\rho}/\sqrt{\rho}$, and accordingly,
the $\mathbb{W}$-gradient flow is the DLSS equation (a.k.a. QDD4),

$$\partial_t \rho = -2\,\mathrm{div}\left(\rho\,\nabla\left[\frac{\Delta\sqrt{\rho}}{\sqrt{\rho}}\right]\right) = -\nabla^2 : \big(\rho\,\nabla^2 \log\rho\big).$$

Fourth order example 2: quantum diffusion

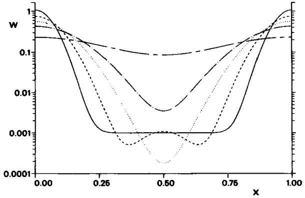
Fisher information

Consider the Fisher information

$$\boldsymbol{F}(\rho) = 2\int \left|\nabla\sqrt{\rho}\right|^2 \mathrm{d}x = \frac{1}{2}\int |\nabla \log\rho|^2 \rho \,\mathrm{d}x.$$

Then $\delta\boldsymbol{E}/\delta\rho = -\Delta\sqrt{\rho}/\sqrt{\rho}$, and accordingly,
the $\mathbb{W}$-gradient flow is the DLSS equation (a.k.a. QDD4),

$$\partial_t\rho = -2\,\mathrm{div}\left(\rho\,\nabla\left[\frac{\Delta\sqrt{\rho}}{\sqrt{\rho}}\right]\right) = -\nabla^2 : \big(\rho\,\nabla^2\log\rho\big).$$



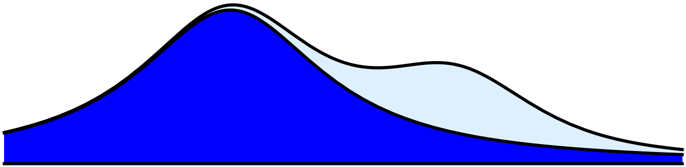
Exotic example 1: Muskat problem

With $a, b > 0$ and $\boldsymbol{\rho} = (\rho_1, \rho_1) \in \mathcal{P}_2(\mathbb{R})^2$, consider

$$\boldsymbol{E}(\boldsymbol{\rho}) = a \int \rho_1^2 \,\mathrm{d}x + b \int (\rho_1 + \rho_2)^2 \,\mathrm{d}x.$$

The gradient flow of $\boldsymbol{E}$ in the $\mathbb{W} \times \mathbb{W}$-metric is the Muskat problem,

$$\partial_t \boldsymbol{\rho} = \partial_x \left(\begin{pmatrix} (a+b)\rho_1 & b\rho_1 \\ \rho_2 & \rho_2 \end{pmatrix} \partial_x \boldsymbol{\rho} \right).$$



Exotic example 2: parabolic–parabolic Keller–Segel

Consider

$$\boldsymbol{E}(\rho, c) = \int \left(\rho \log \rho - \rho c + \frac{1}{2}|\nabla c|^2 + \frac{1}{2}c^2\right) \mathrm{d}x.$$

The gradient flow of $\boldsymbol{E}$ in the $\mathbb{W} \times L^2$-metric is the Keller-Segel system,

$$\begin{aligned} \partial_t \rho &= \Delta \rho - \operatorname{div}(\rho \nabla c) \\ \partial_t c &= \Delta c - c + \rho. \end{aligned}$$

Got it . . . but why should I care?

There are benefits from
being a Wasserstein gradient flow.

Benefit 1: Variational form

Morally, the gradient flow is such that, at any $t > 0$,

$$\partial_t \rho_t = \mathrm{div}(\rho_t \nabla \varphi_t) \quad \text{where} \quad \varphi_t \in \operatorname*{argmin}_{\varphi} \left(\frac{1}{2} \|\varphi\|^2_{\rho_t} + \mathrm{D}_{\rho_t} \boldsymbol{E}[\varphi] \right).$$

Benefit 1: Variational form

Morally, the gradient flow is such that, at any $t > 0$,

$$\partial_t \rho_t = \operatorname{div}(\rho_t \nabla \varphi_t) \quad \text{where} \quad \varphi_t \in \operatorname*{argmin}_{\varphi} \left(\frac{1}{2} \|\varphi\|_{\rho_t}^2 + \mathrm{D}_{\rho_t} \boldsymbol{E}[\varphi] \right).$$

Among many things, this implies Hölder continuity:

Hölder estimate

For $t_2 > t_1 > 0$,

$$\frac{\mathbb{W}(\rho_{t_1}, \rho_{t_2})^2}{t_2 - t_1} \leq \int_{t_1}^{t_2} \|\varphi_t\|_{\rho_t}^2 \, \mathrm{d}t \leq \boldsymbol{E}(\rho_{t_1}) - \boldsymbol{E}(\rho_{t_2}).$$

Benefit 2: Dissipation

Morally, the gradient flow balances kinetic energy and dissipation,

$$\|\varphi_t\|_{\rho_t}^2 = |\partial \boldsymbol{E}(\rho_t)|^2 := -\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{E}(\rho_t) = \int \left|\nabla \frac{\delta \boldsymbol{E}}{\delta \rho}\right|^2 \rho_t \,\mathrm{d}x.$$

This provides a priori estimates:

Dissipation estimate

For $t_2 > t_1 > 0$,

$$\int_{t_1}^{t_2} |\partial \boldsymbol{E}(\rho_t)|^2 \,\mathrm{d}t \leq \boldsymbol{E}(\rho_{t_2}) - \boldsymbol{E}(\rho_{t_1}).$$

Benefit 2: Dissipation

Morally, the gradient flow balances kinetic energy and dissipation,

$$\|\varphi_t\|_{\rho_t}^2 = |\partial \boldsymbol{E}(\rho_t)|^2 := -\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{E}(\rho_t) = \int \left|\nabla \frac{\delta \boldsymbol{E}}{\delta \rho}\right|^2 \rho_t \, \mathrm{d}x.$$

This provides a priori estimates:

Dissipation estimate

For $t_2 > t_1 > 0$,

$$\int_{t_1}^{t_2} |\partial \boldsymbol{E}(\rho_t)|^2 \, \mathrm{d}t \leq \boldsymbol{E}(\rho_{t_2}) - \boldsymbol{E}(\rho_{t_1}).$$

Example: The Boltzmann entropy $\boldsymbol{H}(\rho) = \int \rho \log \rho \, \mathrm{d}x$ has dissipation

$$|\partial \boldsymbol{H}(\rho)|^2 = \int \left|\nabla \log \rho\right|^2 \rho \, \mathrm{d}x = 4 \int |\nabla \sqrt{\rho}|^2 \, \mathrm{d}x = 2\boldsymbol{F}(\rho),$$

i.e., twice the Fisher information. Thus, for $\partial_t \rho = \Delta \rho$,

$$2 \int_{t_1}^{t_2} \boldsymbol{F}(\rho_t) \, \mathrm{d}t \leq \boldsymbol{H}(\rho_{t_1}) - \boldsymbol{H}(\rho_{t_2}).$$

Benefit 3: Positivity

Consider the thin film equation:

$$\partial_t \rho = -\operatorname{div}\big(\rho\, \nabla \Delta \rho\big)$$

From the PDE alone, it is not clear that solutions

- remain non-negative and
- preserve mass.

But the fact that it is formally the $\mathbb{W}$-gradient flow of the Dirichlet energy

$$\boldsymbol{E}(\rho) = \frac{1}{2} \int |\nabla \rho|^2 \,\mathrm{d}x$$

is a strong indication — and actually provides an angle of attack.

There is one more significant thing:

Convexity!

A (fake) ODE analogy for convexity (1/2)

Consider the euclidean gradient flow $\dot{x} = -\nabla \boldsymbol{e}(x)$ for a λ-uniformly convex energy $\boldsymbol{e} : \mathbb{R}^n \to \mathbb{R}$, i.e.,

$$\nabla^2 \boldsymbol{e} \geq \lambda \mathbf{1}.$$

A (fake) ODE analogy for convexity (1/2)

Consider the euclidean gradient flow $\dot{x} = -\nabla \boldsymbol{e}(x)$ for a λ-uniformly convex energy $\boldsymbol{e} : \mathbb{R}^n \to \mathbb{R}$, i.e.,

$$\nabla^2 \boldsymbol{e} \geq \lambda \mathbf{1}.$$

Lemma

If $\lambda > 0$, then:

- $\boldsymbol{e}$ has a unique minimizer $x_\infty \in \mathbb{R}^n$,
- speed, dissipation, energy and distance to x_∞ decay exponentially,

$$\begin{aligned} \left\|\dot{x}_t\right\|^2 &\leq e^{-2\lambda t}\left\|\dot{x}_0\right\|^2, \\ \left\|\nabla \boldsymbol{e}(x_t)\right\|^2 &\leq e^{-2\lambda t}\left\|\nabla \boldsymbol{e}(x_0)\right\|^2, \\ \boldsymbol{e}(x_t) - \boldsymbol{e}(x_\infty) &\leq e^{-2\lambda t}\left(\boldsymbol{e}(x_0) - \boldsymbol{e}(x_\infty)\right), \\ \left\|x_t - x_\infty\right\|^2 &\leq e^{-2\lambda t}\left\|x_0 - x_\infty\right\|^2; \end{aligned}$$

- the flow is λ-uniformly contractive,

$$\left\|\hat{x}_t - \check{x}_t\right\|^2 \leq e^{-2\lambda t}\left\|\hat{x}_0 - \check{x}_0\right\|^2.$$

A (fake) ODE analogy for convexity (2/2)

For $\boldsymbol{e} \in C^2(\mathbb{R}^n)$, λ-uniform convexity is equivalent to

$$\boldsymbol{e}(x_s) \leq (1-s)\boldsymbol{e}(x_0) + s\boldsymbol{e}(x_1) - \frac{\lambda}{2}s(1-s)\|x_1 - x_0\|^2$$

for any choice of $x_0, x_1 \in \mathbb{R}^n$ and $s \in [0,1]$,
where $(x_s)_{s\in[0,1]}$ is the straight line from x_0 to x_1.

A (fake) ODE analogy for convexity (2/2)

For $\boldsymbol{e} \in C^2(\mathbb{R}^n)$, λ-uniform convexity is equivalent to

$$\boldsymbol{e}(x_s) \leq (1-s)\boldsymbol{e}(x_0) + s\boldsymbol{e}(x_1) - \frac{\lambda}{2}s(1-s)\|x_1 - x_0\|^2$$

for any choice of $x_0, x_1 \in \mathbb{R}^n$ and $s \in [0,1]$,
where $(x_s)_{s\in[0,1]}$ is the straight line from x_0 to x_1.

What about simply replacing

- points $x \in \mathbb{R}^n$ by densities $\rho \in \mathcal{P}_2(\mathbb{R}^d)$,
- euclidean norm $\|\cdot\|$ by Wasserstein metric $\mathbb{W}$, and
- straight line $(x_s)_{s\in[0,1]}$ by $\mathbb{W}$-geodesic $(\rho_s)_{s\in[0,1]}$?

Geodesic convexity (1/3)

Definition

$\boldsymbol{E}$ *is* *λ-geodesically convex* *if for any* *$\mathbb{W}$-geodesic* $(\rho_s)_{s\in[0,1]}$*:*

$$\boldsymbol{E}(\rho_s) \leq (1-s)\boldsymbol{E}(\rho_0) + s\boldsymbol{E}(\rho_1) - \frac{\lambda}{2}s(1-s)\mathbb{W}(\rho_0,\rho_1)^2.$$

Geodesic convexity (1/3)

Definition

$\boldsymbol{E}$ is λ-geodesically convex if for any $\mathbb{W}$-geodesic $(\rho_s)_{s\in[0,1]}$:

$$\boldsymbol{E}(\rho_s) \leq (1-s)\boldsymbol{E}(\rho_0) + s\boldsymbol{E}(\rho_1) - \frac{\lambda}{2}s(1-s)\mathbb{W}(\rho_0,\rho_1)^2.$$

Examples:

- the Boltzmann entropy $\boldsymbol{H}$ is 0-geodesically convex;
- if $\nabla^2 V \geq \lambda$, then the potential energy is λ-geodesically convex;
- if $\nabla^2 W \geq \lambda$, then the interaction energy is (almost) λ-geodesically convex;
- if $m > (d-1)/d$, then the internal energy with $h(r) = r^m/(m-1)$ is 0-geodesically convex;
- the Dirichlet energy is not λ-geodesically convex for any λ.

Geodesic convexity (2/3)

Back to the gradient flow $\partial_t \rho = \operatorname{div}\left(\rho \nabla \frac{\delta \boldsymbol{E}}{\delta \rho}\right)$.

Theorem

If $\boldsymbol{E}$ is λ-geodesically convex with $\lambda > 0$, then

- *$\boldsymbol{E}$ possesses precisely one minimizer ρ_∞;*
- *speed, dissipation, energy excess and distance to ρ_∞ decay exponentially,*

$$\begin{aligned}
\big\|\varphi_t\big\|_{\rho_t}^2 &\leq e^{-2\lambda t}\big\|\varphi_0\big\|_{\rho_0}, \\
\big|\partial \boldsymbol{E}(\rho_t)\big|^2 &\leq e^{-2\lambda t}\big|\partial \boldsymbol{E}(\rho_0)\big|^2, \\
\boldsymbol{E}(\rho_t) - \boldsymbol{E}(\rho_\infty) &\leq e^{-2\lambda t}\big(\boldsymbol{E}(\rho_0) - \boldsymbol{E}(\rho_\infty)\big), \\
\mathbb{W}\big(\rho_t, \rho_\infty\big)^2 &\leq e^{-2\lambda t}\mathbb{W}\big(\rho_0, \rho_\infty\big)^2.
\end{aligned}$$

- *the flow is λ-uniformly contractive,*

$$\mathbb{W}\big(\hat{\rho}_t, \check{\rho}_t\big)^2 \leq e^{-2\lambda t}\mathbb{W}\big(\hat{\rho}_0, \check{\rho}_0\big)^2.$$

Geodesic convexity (3/3)

Applications:

- Concerning decay to equilibrium, you will see a truly bombastic application in a couple of slides.
- Concerning contractivity, one obtains e.g. temporal robustness like: Solutions to $\partial_t \rho = \Delta \rho^m$ with $m > \frac{d-1}{d}$ satisfy

$$\mathbb{W}(\rho_{t+\tau}, \rho_t) \leq \mathbb{W}(\rho_\tau, \rho_0).$$

  This is easily seen for $m = 1$, but not trivial for $m \neq 1$.

Geodesic convexity (3/3)

Applications:

- Concerning decay to equilibrium, you will see a truly bombastic application in a couple of slides.
- Concerning contractivity, one obtains e.g. temporal robustness like: Solutions to $\partial_t \rho = \Delta \rho^m$ with $m > \frac{d-1}{d}$ satisfy

$$\mathbb{W}(\rho_{t+\tau}, \rho_t) \leq \mathbb{W}(\rho_\tau, \rho_0).$$

  This is easily seen for $m = 1$, but not trivial for $m \neq 1$.

Proof: would require a more precise definition of $\mathbb{W}$-gradient flow.

As a weak remedy, we obtain contractivity under a sligthly stronger hypothesis.

Displacement convexity (1/2)

Definition

A segment $(\rho_s)_{s\in[0,1]}$ *w.r.t.* η *is such that*

$$\rho_s = \nabla\big((1-s)\varphi_0 + s\varphi_1\big)\#\eta,$$

and φ_0, φ_1 *are optimal, i.e.,*

$$\mathbb{W}(\rho_i,\eta)^2 = \int |\nabla\varphi_i(x) - x|^2\eta(x)\,\mathrm{d}x \quad \textit{for } i = 0, 1.$$

Note: Segments are more general than geodesics.

Displacement convexity (1/2)

Definition

A segment $(\rho_s)_{s\in[0,1]}$ *w.r.t.* η *is such that*

$$\rho_s = \nabla\big((1-s)\varphi_0 + s\varphi_1\big)\#\eta,$$

and φ_0, φ_1 *are optimal, i.e.,*

$$\mathbb{W}(\rho_i,\eta)^2 = \int |\nabla\varphi_i(x) - x|^2\eta(x)\,\mathrm{d}x \quad \text{for } i = 0, 1.$$

Note: Segments are more general than geodesics.

Definition

$\boldsymbol{E}$ *is* λ*-displacement convex if for any segment* $(\rho_s)_{s\in[0,1]}$*:*

$$\boldsymbol{E}(\rho_s) \le (1-s)\boldsymbol{E}(\rho_0) + s\boldsymbol{E}(\rho_1) - \frac{\lambda}{2}s(1-s)\mathbb{W}(\rho_0,\rho_1)^2.$$

Note: λ-displacement convexity implies λ-geodesically convexity. All of the aforementioned λ-geodesically convex functionals happen to be λ-displacement convex as well.

Displacement convexity (2/2)

Definition

$(\rho_t)_{t\geq 0}$ *is an* $(\mathrm{EVI})_\lambda$*-solution if, for all* $\eta \in \mathcal{P}_2(\mathbb{R}^d)$*,*

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{W}(\rho_t,\eta)^2 + \frac{\lambda}{2}\mathbb{W}(\rho_t,\eta)^2 \leq \boldsymbol{E}(\eta) - \boldsymbol{E}(\rho_t).$$

Displacement convexity (2/2)

Definition

$(\rho_t)_{t\geq 0}$ *is an* $(\mathrm{EVI})_\lambda$*-solution if, for all* $\eta \in \mathcal{P}_2(\mathbb{R}^d)$*,*

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{W}(\rho_t,\eta)^2 + \frac{\lambda}{2}\mathbb{W}(\rho_t,\eta)^2 \leq \boldsymbol{E}(\eta) - \boldsymbol{E}(\rho_t).$$

Theorem

If $\boldsymbol{E}$ *is* λ*-geodesically convex,*
then its gradient flow has an $(\mathrm{EVI})_\lambda$*-solution.*

Proof: will be given tomorrow, under the stronger hypothesis of λ-displacement convexity.

Displacement convexity (2/2)

Definition

$(\rho_t)_{t\geq 0}$ *is an* $(\mathrm{EVI})_\lambda$*-solution if, for all* $\eta \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{W}(\rho_t,\eta)^2 + \frac{\lambda}{2}\mathbb{W}(\rho_t,\eta)^2 \leq \boldsymbol{E}(\eta) - \boldsymbol{E}(\rho_t).$$

Theorem

If $\boldsymbol{E}$ *is* λ*-geodesically convex,*
then its gradient flow has an $(\mathrm{EVI})_\lambda$*-solution.*

Proof: will be given tomorrow, under the stronger hypothesis of λ-displacement convexity.

Theorem

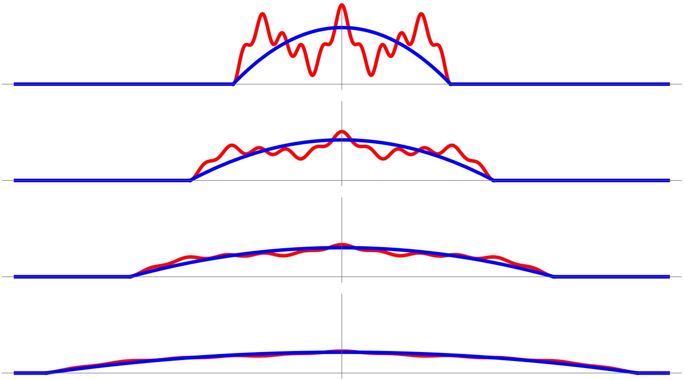
$(\mathrm{EVI})_\lambda$*-solutions are* λ*-contractive.*

...see computations on board ...

Just one more thing!

(It's about atomic bombs!)

A “real” application (1/2)



A “real” application (2/2)

The porous medium equation

$$\partial_t u = \Delta u^m \qquad (\mathsf{PME}_m)$$

has the self-similar solution

$$u_t^*(x) = t^{-d\alpha} U_m(t^{-\alpha} x) \quad \text{with} \quad U_m(z) = \left(B_m - \frac{m-1}{2m} |z|^2 \right)_+^{\frac{1}{m-1}} .$$

A "real" application (2/2)

The porous medium equation

$$\partial_t u = \Delta u^m \qquad (\mathsf{PME}_m)$$

has the self-similar solution

$$u_t^*(x) = t^{-d\alpha} U_m(t^{-\alpha} x) \quad \text{with} \quad U_m(z) = \left(B_m - \frac{m-1}{2m} |z|^2 \right)_+^{\frac{1}{m-1}}.$$

Theorem ([Otto'01])

$$\|u_t - u_t^*\|_{L^1} \leq C t^{-\alpha}.$$

A “real” application (2/2)

The porous medium equation

$$\partial_t u = \Delta u^m \tag{PME$_m$}$$

has the self-similar solution

$$u_t^*(x) = t^{-d\alpha} U_m(t^{-\alpha} x) \quad \text{with} \quad U_m(z) = \left(B_m - \frac{m-1}{2m} |z|^2 \right)_+^{\frac{1}{m-1}}.$$

Theorem ([Otto'01])

$$\|u_t - u_t^*\|_{L^1} \le C t^{-\alpha}.$$

Proof: Use rescaling to rewrite (PME$_m$) as

$$\partial_t \rho = \Delta \rho^m + \operatorname{div}(x\rho),$$

which the $\mathbb{W}$-gradient flow of the geodesically 1-convex functional

$$\boldsymbol{H}'_m(\rho) = \int \left(\frac{\rho^m}{m-1} + \frac{1}{2}|x|^2 \rho \right) \mathrm{d}x.$$

And here is a peaceful spin-off (1/2)

An analogous result holds for the DLSS equation

$$\partial_t u = -2\,\mathrm{div}\left(u\,\nabla\frac{\Delta\sqrt{u}}{\sqrt{u}}\right), \tag{DLSS}$$

which has the self-similar solution

$$u_t^*(x) = t^{-d/4}U_1(t^{-1/4}x) \quad \text{with} \quad U_1(z) = (4\pi)^{-d/2}\exp\left(-\frac{1}{2}|z|^2\right).$$

And here is a peaceful spin-off (1/2)

An analogous result holds for the DLSS equation

$$\partial_t u = -2\,\mathrm{div}\left(u\,\nabla\frac{\Delta\sqrt{u}}{\sqrt{u}}\right), \qquad \text{(DLSS)}$$

which has the self-similar solution

$$u_t^*(x) = t^{-d/4}U_1(t^{-1/4}x) \quad \text{with} \quad U_1(z) = (4\pi)^{-d/2}\exp\left(-\frac{1}{2}|z|^2\right).$$

Theorem ([M'+McCann+Savaré'09])

$$\|u_t - u_t^*\|_{L^1} \leq Ct^{-1/4}.$$

And here is a peaceful spin-off (1/2)

An analogous result holds for the DLSS equation

$$\partial_t u = -2\,\mathrm{div}\left(u\,\nabla\frac{\Delta\sqrt{u}}{\sqrt{u}}\right), \qquad \text{(DLSS)}$$

which has the self-similar solution

$$u_t^*(x) = t^{-d/4}U_1(t^{-1/4}x) \quad \text{with} \quad U_1(z) = (4\pi)^{-d/2}\exp\left(-\frac{1}{2}|z|^2\right).$$

Theorem ([M'+McCann+Savaré'09])

$$\|u_t - u_t^*\|_{L^1} \le Ct^{-1/4}.$$

Proof: Use rescaling to rewrite (DLSS) as

$$\partial_t\rho = -2\,\mathrm{div}\left(\rho\,\nabla\frac{\Delta\sqrt{\rho}}{\sqrt{\rho}}\right) + \mathrm{div}(x\rho),$$

which the $\mathbb{W}$-gradient flow of

$$\boldsymbol{F}'(\rho) = \boldsymbol{F}(\rho) + \frac{1}{2}\int |x|^2\rho\,\mathrm{d}x.$$

And here is a peaceful spin-off (1/2)

An analogous result holds for the DLSS equation

$$\partial_t u = -2\,\mathrm{div}\left(u\,\nabla\frac{\Delta\sqrt{u}}{\sqrt{u}}\right), \qquad \text{(DLSS)}$$

which has the self-similar solution

$$u_t^*(x) = t^{-d/4}U_1(t^{-1/4}x) \quad \text{with} \quad U_1(z) = (4\pi)^{-d/2}\exp\left(-\frac{1}{2}|z|^2\right).$$

Theorem ([M'+McCann+Savaré'09])

$$\|u_t - u_t^*\|_{L^1} \le Ct^{-1/4}.$$

Proof: Use rescaling to rewrite (DLSS) as

$$\partial_t\rho = -2\,\mathrm{div}\left(\rho\,\nabla\frac{\Delta\sqrt{\rho}}{\sqrt{\rho}}\right) + \mathrm{div}(x\rho),$$

which the $\mathbb{W}$-gradient flow of the **totally non**-displacement convex

$$\boldsymbol{F}'(\rho) = \boldsymbol{F}(\rho) + \frac{1}{2}\int |x|^2\rho\,\mathrm{d}x.$$

And here is a peaceful spin-off (2/2)

Theory guanrantees 1-displacement convexity of the relative entropy,

$$\boldsymbol{H}'(\rho) = \int \rho \log \rho \, \mathrm{d}x + \frac{1}{2} \int |x|^2 \rho \, \mathrm{d}x.$$

And here is a peaceful spin-off (2/2)

Theory guanrantees 1-displacement convexity of the relative entropy,

$$\boldsymbol{H}'(\rho) = \int \rho \log \rho \, \mathrm{d}x + \frac{1}{2} \int |x|^2 \rho \, \mathrm{d}x.$$

There is an intimate connection to the relative Fisher information:

$$\begin{aligned}
\left|\partial \boldsymbol{H}'(\rho)\right|^2 &= \int \left|\nabla \log \rho + x\right|^2 \rho \, \mathrm{d}x \\
&= \int \left(|\nabla \log \rho|^2 \rho + |x|^2 \rho + 2 \nabla \rho \cdot x\right) \mathrm{d}x \\
&= 2\boldsymbol{F}(\rho) + \int |x|^2 \rho \, \mathrm{d}x - 2d \int \rho \, \mathrm{d}x \\
&= 2\boldsymbol{F}'(\rho) - 2d.
\end{aligned}$$

And here is a peaceful spin-off (2/2)

Theory guanrantees 1-displacement convexity of the relative entropy,

$$\boldsymbol{H}'(\rho) = \int \rho \log \rho \, \mathrm{d}x + \frac{1}{2} \int |x|^2 \rho \, \mathrm{d}x.$$

There is an intimate connection to the relative Fisher information:

$$\begin{aligned} \left|\partial \boldsymbol{H}'(\rho)\right|^2 &= \int \left|\nabla \log \rho + x\right|^2 \rho \, \mathrm{d}x \\ &= \int \left(|\nabla \log \rho|^2 \rho + |x|^2 \rho + 2 \nabla \rho \cdot x\right) \mathrm{d}x \\ &= 2\boldsymbol{F}(\rho) + \int |x|^2 \rho \, \mathrm{d}x - 2d \int \rho \, \mathrm{d}x \\ &= 2\boldsymbol{F}'(\rho) - 2d. \end{aligned}$$

Theorem

$$\begin{aligned} \boldsymbol{H}'(\rho_t) - \boldsymbol{H}'(\rho_\infty) \leq e^{-2t}\big(\boldsymbol{H}'(\rho_0) - \boldsymbol{H}'(\rho_\infty)\big), \\ \boldsymbol{F}'(\rho_t) - \boldsymbol{F}'(\rho_\infty) \leq e^{-2t}\big(\boldsymbol{F}'(\rho_0) - \boldsymbol{F}'(\rho_\infty)\big). \end{aligned}$$

That's all for today

Thank you, and see you tomorrow!