Application of Optimal Transport to Evolutionary PDEs

4 - Displacement convexity

Giuseppe Savaré
http://www.imati.cnr.it/∼savare
Department of Mathematics, University of Pavia, Italy

2010 CNA Summer School
New Vistas in Image Processing and PDEs

Carnegie Mellon Center for Nonlinear Analysis, Pittsburgh, *June 7–12, 2010*

Outline

Outline

Displacement interpolation

If $\mu_0, \mu_1 \in \mathscr{P}_2(\mathbb{R}^d)$ and $\boldsymbol{\mu} \in \Gamma_o(\mu_0, \mu_1)$ is an optimal coupling, we can consider the **displacement interpolating** curve μ_θ, $\theta \in [0,1]$,

$$\boxed{\mu_\theta = \big((1-\theta)\pi^1 + \theta\pi^2\big)_\# \boldsymbol{\mu}, \quad \pi^1(x_1,x_2) = x_1, \quad \pi^2(x_1,x_2) = x_2}$$

Equivalent probabilistic notation: let X_0, X_1 be an optimal couple of random variables, $(X_i)_\#\mathbb{P} = \mu_i$, $(X_0,X_1)_\#\mathbb{P} = \boldsymbol{\mu}$. $\mu_\theta = \big((1-\theta)X_0 + \theta X_1\big)_\#\mathbb{P}$ is just the law of the interpolated random variable $X_\theta := (1-\theta)X_0 + \theta X_1$

μ_θ is a **minimal, constant speed geodesic** connecting μ_0 and μ_1, since

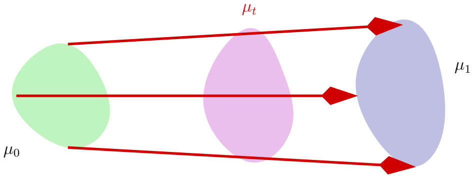
$$W_2(\mu_s, \mu_t) = |t-s| W_2(\mu_0, \mu_1), \quad |\dot\mu_\theta| \equiv W_2(\mu_0,\mu_1) = \mathcal{L}[\mu].$$

When $\mu_0 \ll \mathscr{L}^d$ is absolutely continuous, then Brenier theorem shows that $\boldsymbol{\mu}$ is unique and it is concentrated on the graph of an optimal map $\boldsymbol{t}$,

$$\boldsymbol{\mu} = (\boldsymbol{i} \times \boldsymbol{t})_\#\mu_0, \quad \mu_\theta = \big((1-\theta)\boldsymbol{i} + \theta\boldsymbol{t}\big)_\#\mu_0.$$

The map $\boldsymbol{t}$ is (cyclically monotone), the interpolated maps $\boldsymbol{t}_\theta(x) := (1-\theta)x + \theta\boldsymbol{t}(x)$, $\theta \in [0,1)$ are also **strongly** monotone, since

$$\langle \boldsymbol{t}(x) - \boldsymbol{t}(y), x-y\rangle \ge 0, \quad \langle \boldsymbol{t}_\theta(x) - \boldsymbol{t}_\theta(y), x-y\rangle \ge (1-\theta)|x-y|^2,$$



Displacement convexity

Definition (McCann '97)

A functional $\Phi : \mathscr{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ is **$\boldsymbol{\lambda}$-displacement convex** if for every couple of measures $\mu_0, \mu_1 \in D(\Phi)$ **there exists** a geodesic μ_θ, $\theta \in [0,1]$ connecting them (equivalently, an optimal coupling $\boldsymbol{\mu} \in \Gamma_o(\mu_0, \mu_1)$) such that

$$\boxed{\Phi(\mu_\theta) \leq (1-\theta)\Phi(\mu_0) + \theta\Phi(\mu_1) - \frac{\boldsymbol{\lambda}}{2}\theta(1-\theta)W_2^2(\mu_0, \mu_1).} \tag{DC}$$

A few comments:

- (DC) is modelled on the analogous inequality for $\boldsymbol{\lambda}$-convex functions in $\mathbb{R}^m$:

$$\phi((1-\theta)x_0 + \theta x_1) \leq (1-\theta)x_0 + \theta x_1 - \frac{\boldsymbol{\lambda}}{2}\theta(1-\theta)\,|x_0 - x_1|^2.$$

  Just replace segments with geodesics and the euclidean distance with the Wasserstein one.

- Since geodesics are not unique, the present definition only requires that convexity inequality holds **at least along one geodesic**. This is sufficient for the applications and enjoys nice stability properties with respect to perturbation of the functional (e.g. under Γ-convergence).

Displacement convexity for potential energy

Let $V : \mathbb{R}^d \to (-\infty, +\infty]$ be a lower semicontinuous potential with the associated potential energy

$$\mathcal{V}(\mu) := \int_{\mathbb{R}^d} V(x)\,\mathrm{d}\mu(x).$$

Notice that the functional $\mathcal{V}$ is linear (thus convex) with respect to the usual linear structure in the space of measures.

Theorem

$\mathcal{V}$ is displacement $\boldsymbol{\lambda}$-convex iff V is $\boldsymbol{\lambda}$-convex.

Proof. Let (X_0, X_1) be an optimal couple of random variables with law μ_0, μ_1. $\mu_\theta = (X_\theta)_\# \mathbb{P}$ where $X_\theta = (1-\theta)X_0 + \theta X_1$ and

$$\begin{aligned}
\mathcal{V}(\mu_\theta) &= \int_{\mathbb{R}^d} V(x)\,\mathrm{d}\mu_\theta = \mathbb{E}\Big[V\big(X_\theta\big)\Big] \le \mathbb{E}\Big[(1-\theta)V(X_0) + \theta V(X_1) - \frac{\boldsymbol{\lambda}}{2}\theta(1-\theta)|X_0 - X_1|^2\Big] \\
&= (1-\theta)\mathbb{E}\Big[V(X_0)\Big] + \theta\mathbb{E}\Big[V(X_1)\Big] - \frac{\boldsymbol{\lambda}}{2}\theta(1-\theta)\mathbb{E}\Big[|X_0 - X_1|^2\Big] \\
&= (1-\theta)\mathcal{V}(\mu_0) + \theta\mathcal{V}(\mu_1) - \frac{\boldsymbol{\lambda}}{2}\theta(1-\theta)W_2^2(\mu_0, \mu_1).
\end{aligned}$$

In the last identity we used the optimality of X_0, X_1, i.e.

$$\mathbb{E}\Big[|X_0 - X_1|^2\Big] = W_2^2(\mu_0, \mu_1).$$

Displacement convexity for the interaction energy

Let $W : \mathbb{R}^d \to (-\infty, +\infty]$ be a lower semicontinuous, even interaction potential with the associated interaction energy

$$\mathcal{W}(\mu) := \frac{1}{2} \iint_{\mathbb{R}^d \times \mathbb{R}^d} W(x-y) \,\mathrm{d}\mu(x) \,\mathrm{d}\mu(y).$$

Theorem

Suppose $\boldsymbol{\lambda} \le 0$: $\mathcal{W}$ is displacement $\boldsymbol{\lambda}$-convex iff W is $\boldsymbol{\lambda}$-convex.

Comments.

- The functional $\mathcal{W}$ is quadratic but it is generally not convex with respect to the usual linear structure in the space of measures. Consider e.g. $W(x) := \frac{1}{2}|x|^2$:

  $$\mathcal{W}(\mu) = \frac{1}{2} \iint |x-y|^2 \,\mathrm{d}\mu(x) \,\mathrm{d}\mu(y) = \int_{\mathbb{R}^d} |x|^2 \,\mathrm{d}\mu(x) - \Big| \int_{\mathbb{R}^d} x \,\mathrm{d}\mu(x) \Big|^2$$

  so that $\mathcal{W}((1-\theta)\delta_0 + \theta\delta_x) = (\theta - \theta^2)|x|^2$
- When W is $\boldsymbol{\lambda}$-covex with $\boldsymbol{\lambda} > 0$ we can only deduce that $\mathcal{W}$ is displacement convex: the functional is in fact invariant by translation.

Displacement convexity for internal energy functionals

Let $F : [0, +\infty) \to \mathbb{R}$ be a continuous convex function with $F(0) = 0$ and let

$$\mathcal{F}(\mu) := \int_{\mathbb{R}^d} F(u(x))\,\mathrm{d}x, \qquad \mu = u\mathscr{L}^d, \quad u = \frac{\mathrm{d}\mu}{\mathrm{d}\mathscr{L}^d}.$$

Theorem (McCann '97)

$\mathcal{F}$ is displacement convex iff $r \mapsto r^{-d}F(r^d)$ is convex and non increasing in $(0, +\infty)$.

Comments:

- When $d = 1$ McCann condition is equivalent to the convexity of F. When $d > 1$ it is stronger.
- The logarithmic entropy $\mathcal{H}$ corresponding to $F(r) = r\log r$ is always displacement convex.
- The power energy $F(r) = \dfrac{1}{\beta - 1} r^\beta$ is displacement convex iff $\beta \geq 1 - 1/d$.

Application: a simple proof of Talagrand inequality

Let us consider a reference probability measure $\gamma = \mathrm{e}^{-V}$ where V satisfies the Bakry-Emery condition $\mathrm{D}^2 V \geq \boldsymbol{\lambda} I$, i.e. it is $\boldsymbol{\lambda}$-convex with $\boldsymbol{\lambda} > 0$.

The typical example is the Gaussian measure $\gamma := (2\pi)^{-d/2}\mathrm{e}^{-|x|^2/2}\,\mathscr{L}^d$, $V(x) := \frac{1}{2}|x|^2 + \frac{d}{2}\log(2\pi)$.
The **relative entropy functional** w.r.t. γ

$$\boxed{\mathcal{H}(\mu|\gamma) := \int_{\mathbb{R}^d} \frac{\mathrm{d}\mu}{\mathrm{d}\gamma}\log\Big(\frac{\mathrm{d}\mu}{\mathrm{d}\gamma}\Big)\,\mathrm{d}\gamma = \mathcal{H}(\mu) + \mathcal{V}(\mu) \quad \text{is displacement } \boldsymbol{\lambda}\text{-convex.}}$$

Jensen inequality yields

$$\mathcal{H}(\mu|\gamma) \geq \mu(\mathbb{R}^d)\log(\mu(\mathbb{R}^d)) = 0 = \mathcal{H}(\gamma|\gamma).$$

so that γ is the (unique) minimizer of $\mathcal{H}(\cdot|\gamma)$ in $\mathscr{P}_2(\mathbb{R}^d)$.

Theorem (Talagrand inequality (Otto-Villani '00))

$$\frac{\boldsymbol{\lambda}}{2}W_2^2(\mu,\gamma) \leq \mathcal{H}(\mu|\gamma) \quad \textit{for every } \mu \in \mathscr{P}_2(\mathbb{R}^d).$$

Proof. Take a geodesic μ_θ conecting γ to μ and apply the convexity inequality

$$\begin{aligned} 0 \leq \mathcal{H}(\mu_\theta|\gamma) &\leq (1-\theta)\mathcal{H}(\gamma|\gamma) + \theta\mathcal{H}(\mu|\gamma) - \frac{\boldsymbol{\lambda}}{2}\theta(1-\theta)W_2^2(\mu,\gamma) \\ &= \theta\mathcal{H}(\mu|\gamma) - \frac{\boldsymbol{\lambda}}{2}\theta(1-\theta)W_2^2(\mu,\gamma). \end{aligned}$$

Dividing by θ and letting $\theta \downarrow 0$ we conclude.

Displacement convexity of the logarithmic entropy

Let $\mu_i = u_i \mathscr{L}^d$ be probability measures with finite entropy, let $\boldsymbol{t}$ the Brenier map pushing μ_0 to μ_1, and let $\mu_\theta = u_\theta \mathscr{L}^d = (\boldsymbol{t}_\theta)_\# \mu_0$ with $\boldsymbol{t}_\theta = (1-\theta)\boldsymbol{i} + \theta \boldsymbol{t}$.
We have already shown that

$$\mathcal{H}(\mu_\theta) = \mathcal{H}(\mu_0) - \int_{\mathbb{R}^d} \log\Big(\det \mathrm{D}\boldsymbol{t}_\theta(x)\Big) u_0(x)\,\mathrm{d}x$$

Thus it is sufficient to show that

$$\boxed{\text{the map} \quad \theta \mapsto \log\Big(\det \mathrm{D}\boldsymbol{t}_\theta(x)\Big) \quad \text{is concave for } \mu_0\text{-a.e. } x \in \mathbb{R}^d.}$$

Notice that $\mathrm{D}\boldsymbol{t}_\theta(x) = (1-\theta)\mathsf{I} + \theta \mathsf{D}(x)$ where I is the identity matrix and $\mathsf{D}(x) = \mathrm{D}\boldsymbol{t}(x)$ **is a symmetric and positive definite matrix** for μ_0-a.e. x (Brenier theorem). Denoting by $\lambda_i(x)$ its positive eigenvalues, we have

$$\log\Big(\det \mathrm{D}\boldsymbol{t}_\theta(x)\Big) = \log\Big(\Pi_{i=1}^d \big(1-\theta+\theta\lambda_i(x)\big)\Big) = \sum_{i=1}^d \log\Big(\big(1-\theta+\theta\lambda_i(x)\big)\Big)$$

which is clearly concave w.r.t. θ.

Outline

1 Displacement convexity

2 Slope and subgradient inequalities: applications to Sobolev and logarithmic Sobolev inequalities

3 Generation results for gradient flows of displacement λ-convex functionals

4 Convergence of the variational scheme by energy identity

Metric slope

Let us recall that the metric slope of a functional $\Phi : \mathscr{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ is defined as

$$\boxed{|\partial\Phi|(\mu) := \limsup_{\sigma\to\mu} \frac{\big(\Phi(\mu) - \Phi(\sigma)\big)_+}{W_2(\sigma,\mu)}.}$$

The slope provides a simple upper bound of the norm of the (weak) Wasserstein gradient of Φ:

$$\|\partial\Phi(\mu)\|_{L^2(\mu;\mathbb{R}^d)} \le |\partial\Phi|(\mu).$$

In fact, if $\boldsymbol{g} = \partial\Phi(\mu)$, $\boldsymbol{\xi} \in \mathrm{C}^\infty_\mathrm{c}(\mathbb{R}^d;\mathbb{R}^d)$, $\dot X = \boldsymbol{\xi}(X)$, and $\mu_\varepsilon := (X_\varepsilon)_\#\mu$, we have by definition

$$\begin{aligned}\int_{\mathbb{R}^d} \langle \boldsymbol{g}, \boldsymbol{\xi}\rangle \,\mathrm{d}\mu &= \lim_{\varepsilon\downarrow 0} \varepsilon^{-1}\Big(\Phi(\mu) - \Phi((X_\varepsilon)_\#\mu)\Big)\\ &\le \limsup_{\varepsilon\downarrow 0} \frac{\big(\Phi(\mu) - \Phi((X_\varepsilon)_\#\mu)\big)_+}{W_2(\mu_\varepsilon,\mu)}\,\frac{W_2(\mu_\varepsilon,\mu)}{\varepsilon}\\ &\le |\partial\Phi|(\mu)\limsup_{\varepsilon\downarrow 0}\frac{W_2(\mu_\varepsilon,\mu)}{\varepsilon} \le |\partial\Phi|(\mu)\|\boldsymbol{\xi}\|_{L^2(\mu;\mathbb{R}^d)}\end{aligned}$$

since

$$\varepsilon^{-2}W_2^2(\mu_\varepsilon,\mu) \le \varepsilon^{-2}\int_{\mathbb{R}^d} |X_\varepsilon(x) - x|^2\,\mathrm{d}\mu \xrightarrow{\varepsilon\downarrow 0} \int_{\mathbb{R}^d} |\boldsymbol{\xi}|^2\,\mathrm{d}\mu.$$

Extremal properties of the slope

Theorem

If Φ is displacement convex then $|\partial\Phi|(\mu) := \boxed{\sup_{\sigma\neq\mu}} \dfrac{\big(\Phi(\mu)-\Phi(\sigma)\big)_+}{W_2(\sigma,\mu)}$.

In particular, if Φ is lower semicontinuous, the map $\mu \mapsto |\partial\Phi|(\mu)$ is also lower semicontinuous in $\mathscr{P}_2(\mathbb{R}^d)$.

Proof. If σ_θ is a geodesic connecting μ with σ, displacement convexity yields

$$\theta \mapsto \frac{\Phi(\sigma_\theta)-\Phi(\mu)}{W_2(\mu,\sigma_\theta)} \quad \text{is nondecreasing, so that}$$

$$\frac{\big(\Phi(\mu)-\Phi(\sigma)\big)_+}{W_2(\sigma,\mu)} \le \frac{\big(\Phi(\mu)-\Phi(\sigma_\theta)\big)_+}{W_2(\sigma_\theta,\mu)} \le \limsup_{\theta\downarrow 0} \frac{\big(\Phi(\mu)-\Phi(\sigma_\theta)\big)_+}{W_2(\sigma_\theta,\mu)} \le |\partial\Phi|(\mu).$$

The "sup" formula for the slope can be equivalently stated as

$$-|\partial\Phi|(\sigma)W_2(\mu,\sigma) \le \Phi(\mu)-\Phi(\sigma) \le |\partial\Phi|(\mu)W_2(\mu,\sigma).$$

When Φ is $\boldsymbol{\lambda}$ convex we have more generally

$$-|\partial\Phi|(\sigma)W_2(\mu,\sigma) + \frac{\boldsymbol{\lambda}}{2}W_2^2(\mu,\sigma) \le \Phi(\mu)-\Phi(\sigma) \le |\partial\Phi|(\mu)W_2(\mu,\sigma) + \frac{\boldsymbol{\lambda}}{2}W_2^2(\mu,\sigma)$$

If $\boldsymbol{\lambda} > 0$ and choosing $\sigma := \mu_{min}$ the minimizer of Φ, we have $|\partial\Phi|(\mu_{min}) = 0$ and

$$\frac{\boldsymbol{\lambda}}{2}W_2^2(\mu,\mu_{min}) \le \Phi(\mu)-\Phi(\mu_{min}) \le \frac{1}{2\boldsymbol{\lambda}}|\partial\Phi|^2(\mu).$$

Strong Wasserstein gradients

Let us suppose that Φ is displacement convex.

Theorem (Strong Wasserstein subgradient)

If $|\partial\Phi|(\mu) < +\infty$ *then there exists a vector field* $\boldsymbol{g} = \partial^\circ\Phi(\mu) \in L^2(\mu;\mathbb{R}^d)$ *such that*

$$\|\boldsymbol{g}\|_{L^2(\mu;\mathbb{R}^d)} = |\partial\Phi|(\mu), \qquad \Phi(\boldsymbol{t}_\#\mu) - \Phi(\mu) \ge \int_{\mathbb{R}^d} \langle \boldsymbol{g}, \boldsymbol{t} - \boldsymbol{i}\rangle \, \mathrm{d}\mu(x) + \mathscr{R}(\boldsymbol{t}),$$

where $\mathscr{R}(\boldsymbol{t}) = o(\|\boldsymbol{t} - \boldsymbol{i}\|_{L^2(\mu;\mathbb{R}^d)})$ *and* $\mathscr{R}(\boldsymbol{t}) = 0$ *if* $\boldsymbol{t}$ *is an optimal map.*

- If Φ is differentiable along smooth vector fields, then a strong gradient is also a weak one.
- Conversely, in all the previous examples concerning potential, interaction, and internal energies (under further suitable conditions...), weak Wasserstein gradients are also strong. They can then be used to obtain **quantitative** estimates on Φ, e.g. in the case of displacement $\boldsymbol{\lambda}$-convex functionals, thus satisfying the metric inequality

$$\frac{\boldsymbol{\lambda}}{2} W_2^2(\mu, \mu_{min}) \le \Phi(\mu) - \Phi(\mu_{min}) \le \frac{1}{2\boldsymbol{\lambda}} |\partial\Phi|^2(\mu).$$

A general result

Let $\Phi(\mu) = \mathcal{V}(\mu) + \mathcal{W}(\mu) + \mathcal{F}(\mu)$ where

- $\mathcal{V}(\mu) = \int_{\mathbb{R}^d} V(x)\,\mathrm{d}\mu(x)$ for a lower semicontinuous and convex potential $V : \mathbb{R}^d \to (-\infty, +\infty]$ with $\Omega \subset D(V) \subset \overline{\Omega}$, Ω open in $\mathbb{R}^d$.
- $\mathcal{W}(\mu) := \int_{\mathbb{R}^d} W(x-y)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y)$ for a $\boldsymbol{\lambda}$-convex and differentiable potential $W : \mathbb{R}^d \to \mathbb{R}$ satisfying a doubling condition.
- $\mathcal{F}(\mu) := \int_{\mathbb{R}^d} F(u)\,\mathrm{d}x$ for a convex and superlinear function $F : [0, +\infty) \to \mathbb{R}$, $F(0) = 0$, satisfying McCann condition.

Theorem (Ambrosio-Gigli-S.)

The weak Wasserstein differential of Φ is characterized by

$$\boldsymbol{g} = \partial\Phi(\mu) \in L^2(\mu;\mathbb{R}^d) \quad \Leftrightarrow \quad \begin{cases} \mu = u\mathscr{L}^d, \quad L(u) \in W^{1,1}_{\mathrm{loc}}(\Omega), \\ u\boldsymbol{g} = \nabla L(u) + u(\nabla V + \nabla W * u) \end{cases}$$

*and it is also a **strong differential**; in particular its $L^2(\mu;\mathbb{R}^d)$ coincides with the metric slope $|\partial\Phi|(\mu)$ of Φ.*

Application: Logarithmic Sobolev inequalities

Let $\gamma = \mathrm{e}^{-V}$ be a reference probability measure where V satisfies the Bakry-Emery condition $\mathrm{D}^2 V \geq \boldsymbol{\lambda} I$, i.e. it is $\boldsymbol{\lambda}$-convex with $\boldsymbol{\lambda} > 0$.
We choose $\Phi(\mu) := \mathcal{H}(\mu|\gamma) = \mathcal{H}(\mu) + \mathcal{V}(\mu)$; the Wasserstein gradient of Φ is, at least formally,

$$\partial\Phi(\mu) = \partial\mathcal{H}(\mu) + \partial\mathcal{V}(\mu) = \frac{\nabla u}{u} + \nabla V = \nabla\Big(\log u + V\Big) = \nabla \log\left(u/\mathrm{e}^{-V}\right) = \nabla \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\gamma}\right).$$

The **relative Fisher information** is the L^2-norm of the Wasserstein gradient

$$\mathcal{I}(\mu|\gamma) := \int_{\mathbb{R}^d} \left|\nabla \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\gamma}\right)\right|^2 \mathrm{d}\mu = \int_{\mathbb{R}^d} \frac{|\nabla u + u\nabla V|^2}{u}\,\mathrm{d}\gamma.$$

Theorem (Logarithmic Sobolev inequality (Gross, $\cdots$, Otto-Villani))

The relative Fisher information is the (squared) slope of the relative entropy and

$$\mathcal{H}(\mu|\gamma) \leq \frac{1}{2\lambda}\mathcal{I}(\mu|\gamma). \tag{LS}$$

The *proof* of (LS) follows from the general inequality for $\boldsymbol{\lambda}$-convex functionals

$$\Phi(\mu) - \Phi(u_{min}) \leq \frac{1}{2\lambda}|\partial\Phi|^2(\mu)$$

once we know that $\mathcal{I}(\mu, \gamma) = |\partial\mathcal{H}(\cdot|\gamma)|^2(\mu)$.

Application: optimal constant in Sobolev inequality

Let $d \geq 3$ and $p := 2^*$ such that $\frac{1}{p} = \frac{1}{2} - \frac{1}{d}$. Then

$$\boxed{\min\Big\{\int_{\mathbb{R}^d} |\nabla w|^2 \,\mathrm{d}x : \int_{\mathbb{R}^d} w^p \,\mathrm{d}x = 1\Big\} = \int_{\mathbb{R}^d} |\nabla w_b|^2 \,\mathrm{d}x,} \quad w_b := \Big(\mathsf{a} + \frac{1}{2(d-1)}|x|^2\Big)^{-d/p}.$$

Proof. Let $w^p =: u$, $w_b^p =: u_b$ and take $F(u) := -d\, u^{1-1/d}$, $L(u) := u^{1-1/d}$,

$$\Phi(\mu) := \mathcal{F}(\mu) + \frac{1}{2}\int_{\mathbb{R}^d} |x|^2 \,\mathrm{d}\mu, \quad \partial\Phi(\mu) = \frac{\nabla L(u)}{u} + x, \quad \Phi(\mu) \geq \Phi(\mu_b).$$

$$|\partial\Phi|^2(\mu) = \int_{\mathbb{R}^d} \Big|\frac{\nabla L(u)}{u} + x\Big|^2 u \,\mathrm{d}x = \int_{\mathbb{R}^d} \frac{|\nabla L(u)|^2}{u} \,\mathrm{d}x + \int_{\mathbb{R}^d} |x|^2 \,\mathrm{d}\mu + 2\int_{\mathbb{R}^d} \langle \nabla L(u), x\rangle \,\mathrm{d}x$$
$$= \int_{\mathbb{R}^d} \frac{|\nabla L(u)|^2}{u} \,\mathrm{d}x + \int_{\mathbb{R}^d} |x|^2 \,\mathrm{d}\mu - 2d\int_{\mathbb{R}^d} L(u) = 2c_d \int_{\mathbb{R}^d} \Big|\nabla u^{1/p}\Big|^2 \,\mathrm{d}x + 2\Phi(\mu),$$

where $c_d := 2\Big(\frac{d-1}{d-2}\Big)^2$. Since $|\partial\Phi|(\mu_b) = 0$ we get

$$c_d \int_{\mathbb{R}^d} \Big|\nabla u_b^{1/p}\Big|^2 \,\mathrm{d}x = -\Phi(\mu_b)$$

$$\Phi(\mu) - \Phi(\mu_b) \leq \frac{1}{2}|\partial\Phi|^2(\mu) = c_d \int_{\mathbb{R}^d} \Big|\nabla u^{1/p}\Big|^2 \,\mathrm{d}x + \Phi(\mu),$$

so that

$$\int_{\mathbb{R}^d} \Big|\nabla u^{1/p}\Big|^2 \geq \int_{\mathbb{R}^d} \Big|\nabla u_b^{1/p}\Big|^2.$$

Outline

Equivalent definitions

Theorem (Ambrosio-Gigli-S.)

Let $\Phi : \mathscr{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ be a proper, lower semicontinuous, and displacement $\boldsymbol{\lambda}$-convex functional and let $\mu : (0, +\infty) \to \mathscr{P}_2(\mathbb{R}^d)$ be a locally Lipschitz curve. The following properties are equivalent:

- *μ satisfies the continuity equation*

$$\partial_t \mu + \mathsf{div}(\mu \boldsymbol{v}) = 0 \qquad \int_{t_0}^{t_1} \int_{\mathbb{R}^d} |\boldsymbol{v}_t|^2 \,\mathrm{d}\mu_t \,\mathrm{d}t < +\infty \quad \textit{for every } 0 < t_0 < t_1 < +\infty$$

  *and **$\boldsymbol{v}$ is a strong Wasserstein subgradient** $\boldsymbol{v}_t = -\partial^\circ \Phi(\mu_t)$ for a.e. t.*

- *μ satisfies the **Evolution Variational inequality** for a.e. $t > 0$*

$$\frac{\mathrm{d}}{\mathrm{d}t} W_2^2(\mu_t, \sigma) \le \Phi(\sigma) - \Phi(\mu_t) - \frac{\boldsymbol{\lambda}}{2} W_2^2(\mu_t, \sigma) \quad \textit{for every } \sigma \in D(\Phi) \qquad \text{(EVI)}$$

- *μ satisfies the **Energy identity***

$$-\frac{\mathrm{d}}{\mathrm{d}t} \Phi(\mu_t) = \frac{1}{2} |\dot\mu_t|^2 + \frac{1}{2} |\partial \Phi|^2(\mu_t) \quad \textit{a.e. in } (0, +\infty),$$

  also in the weaker integrated–inequality form (when $\mu_0 \in D(\Phi)$)

$$\Phi(\mu_t) + \frac{1}{2} \int_0^t \Big(|\dot\mu_r|^2 + |\partial \Phi|^2(\mu_r) \Big) \,\mathrm{d}r \le \Phi(\mu_0).$$

A key ingredient: the chain rule

Theorem (Chain rule)

Let $\Phi : \mathscr{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ be a proper, lower semicontinuous and $\boldsymbol{\lambda}$-convex functional, and let $\mu : [a,b] \to \mathbb{R}^d$ be a Lipschitz curve.
If $t \mapsto |\partial\Phi|(\mu_t)$ is integrable in (a,b) then the map $t \mapsto \Phi(\mu_t)$ is absolutely continuous and

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi(\mu_t) = \int_{\mathbb{R}^d} \langle \boldsymbol{v}_t, \partial^\circ\Phi(\mu_t)\rangle \,\mathrm{d}\mu_t \geq -|\dot\mu_t|\,|\partial\Phi|(\mu_t) \quad \textit{a.e. in } (a,b),$$

where $\boldsymbol{v}_t$ is the Wasserstein velocity of μ.

If μ satisfies the energy inequality then

$$\frac{1}{2}\int_0^t \Big(|\dot\mu_r|^2 + |\partial\Phi|^2(\mu_r)\Big)\,\mathrm{d}r \leq \Phi(\mu_0) - \Phi(\mu_t) \quad \stackrel{\textit{chain rule}}{=} -\int_0^t \int_{\mathbb{R}^d} \langle \boldsymbol{v}_r, \partial^\circ\Phi(\mu_r)\rangle \,\mathrm{d}\mu_r$$

so that

$$\frac{1}{2}\int_0^t \Big(|\dot\mu_r|^2 + |\partial\Phi|^2(\mu_r) + 2\int_{\mathbb{R}^d} \langle \boldsymbol{v}_r, \partial^\circ\Phi(\mu_r)\rangle \,\mathrm{d}\mu_r\Big)\,\mathrm{d}r \leq 0$$

i.e.

$$\boldsymbol{v}_t = -\partial^\circ\Phi(\mu_t) \quad \text{for a.e. } t > 0.$$

The main generation result

Theorem (Ambrosio-Gigli-S. 05, S. 07)

Let $\Phi : \mathscr{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ be a proper, lower semicontinuous, and displacement $\boldsymbol{\lambda}$-convex functional. For every $\mu_0 \in \overline{D(\Phi)}$ there exists a unique curve $\mu_t = \mathsf{S}_t[\mu_0]$ solution of (EVI) such that $\lim_{t\downarrow 0} \mu_t = \mu_0$.

- *The map $t \mapsto \mathsf{S}_t[\cdot]$ is a* ***continuous semigroup of $\boldsymbol{\lambda}$ contractions*** *in $\overline{D(\Phi)} \subset \mathscr{P}_2(\mathbb{R}^d)$,*

$$W_2(\mathsf{S}_t[\mu_0], \mathsf{S}_t[\mu_1]) \le \mathrm{e}^{-\boldsymbol{\lambda} t} W_2(\mu_0, \mu_1)$$

- *μ is* ***locally Lipschitz*** *in $(0, +\infty)$, for every $t > 0$, $\mathsf{S}_t[\mu_0] \in D(\partial\Phi) \subset D(\Phi)$, and satisfies the* ***regularization estimate*** *(here $\boldsymbol{\lambda} = 0$)*

$$\frac{1}{2} W_2^2(\mu_t, \sigma) + t\big(\Phi(\mu_t) - \Phi(\sigma)\big) + \frac{t^2}{2} |\partial\Phi|^2(\mu_t) \le \frac{1}{2} W_2^2(\mu_0, \sigma) \quad \forall\, \sigma \in D(\Phi)$$

- *The curves $t \mapsto \mu_t$ and $t \mapsto \Phi(\mu_t)$ are* ***right differentiable*** *at every $t > 0$ and satisfies the* ***minimal selection principle***

$$t \mapsto -\frac{\mathrm{d}}{\mathrm{d}t_+} \Phi(\mu_t) = |\dot\mu_{t+}|^2 = |\partial\Phi|^2(\mu_t) \quad \textit{is nonincreasing.}$$

- ***Asymptotic decay, $\boldsymbol{\lambda} > 0$*** $\quad W_2(\mu_t, \mu_{\min}) \le \mathrm{e}^{-\boldsymbol{\lambda} t} W_2(\mu_0, \mu_{\min})$

$$\Phi(\mu_t) - \Phi(\mu_{\min}) \le \mathrm{e}^{-2\boldsymbol{\lambda} t}\Big(\Phi(\mu_0) - \Phi(\mu_{\min})\Big), \quad |\partial\Phi|^2(\mu_t) \quad \le \mathrm{e}^{-2\boldsymbol{\lambda} t} |\partial\Phi|^2(\mu_0)$$

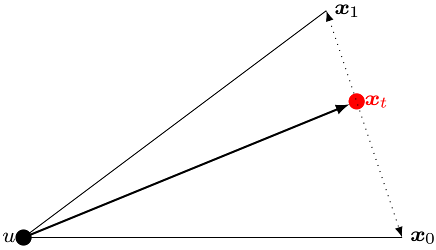
Outline

Convexity properties of the distance

The squared distance enters in a crucial way in the minimizing functional

$$\boldsymbol{M} \mapsto \frac{1}{2\boldsymbol{\tau}} W_2^2(M_\tau^{n-1}, \boldsymbol{M}) + \Phi(\boldsymbol{M}).$$

The **behaviour of the squared distance along geodesics** should play a crucial role. In the Euclidean case



$$\mathsf{d}^2(u, \boldsymbol{x}_{\boldsymbol{\theta}}) = (1-\boldsymbol{\theta})\mathsf{d}^2(u, \boldsymbol{x}_0) + \boldsymbol{\theta}\mathsf{d}^2(u, \boldsymbol{x}_1) - \boldsymbol{\theta(1-\theta)}\mathsf{d}^2(\boldsymbol{x}_0, \boldsymbol{x}_1).$$

The Wasserstein space is Positively Curved (PC)

In $\mathbb{R}^m = \mathscr{P}_2(\mathbb{R}^2)$ consider two point masses $\boldsymbol{\mu_0}$ and $\boldsymbol{\mu_1}$ ...

The Wasserstein space is Positively Curved (PC)

In $\mathbb{R}^m = \mathscr{P}_2(\mathbb{R}^2)$ consider two point masses $\boldsymbol{\mu_0}$ and $\boldsymbol{\mu_1}$ ...

The Wasserstein space is Positively Curved (PC)

In $\mathbb{R}^m = \mathscr{P}_2(\mathbb{R}^2)$ consider two point masses $\boldsymbol{\mu_0}$ and $\boldsymbol{\mu_1}$ ...

The Wasserstein space is Positively Curved (PC)

In $\mathbb{R}^m = \mathscr{P}_2(\mathbb{R}^2)$ consider two point masses $\boldsymbol{\mu_0}$ and $\boldsymbol{\mu_1}$ ...

The Wasserstein space is Positively Curved (PC)

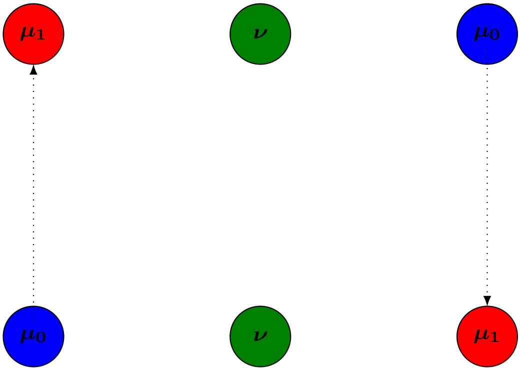
In $\mathbb{R}^m = \mathscr{P}_2(\mathbb{R}^2)$ consider two point masses $\boldsymbol{\mu_0}$ and $\boldsymbol{\mu_1}$ ...

The Wasserstein space is Positively Curved (PC)

In $\mathbb{R}^m = \mathscr{P}_2(\mathbb{R}^2)$ consider two point masses $\boldsymbol{\mu_0}$ and $\boldsymbol{\mu_1}$ ...

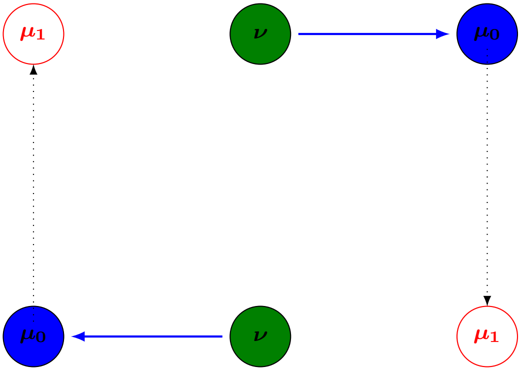
The Wasserstein space is Positively Curved (PC)

In $\mathbb{R}^m = \mathscr{P}_2(\mathbb{R}^2)$ consider two point masses $\boldsymbol{\mu_0}$ and $\boldsymbol{\mu_1}$ ...and a third reference measure $\boldsymbol{\nu}$.



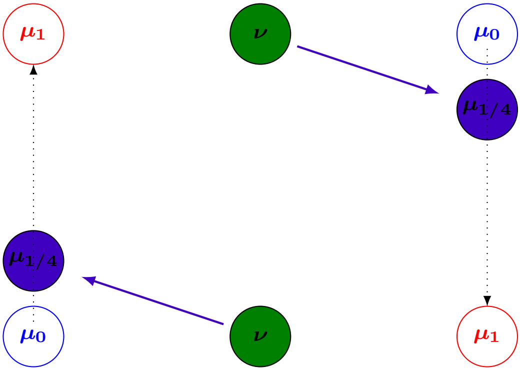
The Wasserstein space is Positively Curved (PC)

In $\mathbb{R}^m = \mathscr{P}_2(\mathbb{R}^2)$ consider two point masses $\boldsymbol{\mu_0}$ and $\boldsymbol{\mu_1}$ …and a third reference measure $\boldsymbol{\nu}$.



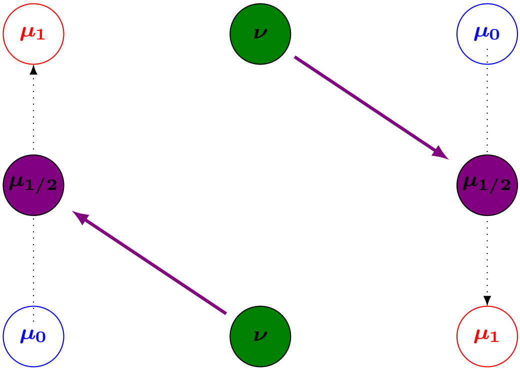
The Wasserstein space is Positively Curved (PC)

In $\mathbb{R}^m = \mathscr{P}_2(\mathbb{R}^2)$ consider two point masses $\boldsymbol{\mu_0}$ and $\boldsymbol{\mu_1}$ ... and a third reference measure $\boldsymbol{\nu}$.



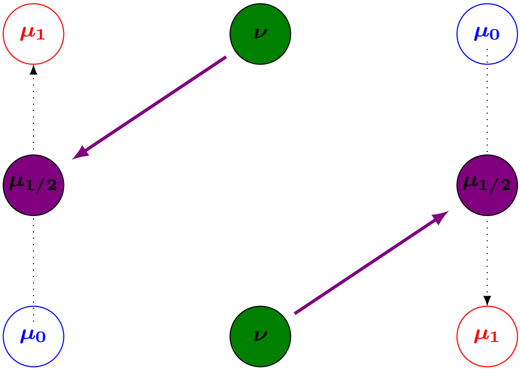
The Wasserstein space is Positively Curved (PC)

In $\mathbb{R}^m = \mathscr{P}_2(\mathbb{R}^2)$ consider two point masses $\boldsymbol{\mu_0}$ and $\boldsymbol{\mu_1}$ ...and a third reference measure $\boldsymbol{\nu}$.



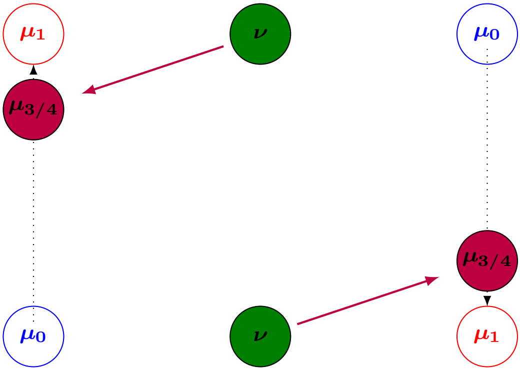
The Wasserstein space is Positively Curved (PC)

In $\mathbb{R}^m = \mathscr{P}_2(\mathbb{R}^2)$ consider two point masses μ_0 and μ_1 ... and a third reference measure ν.



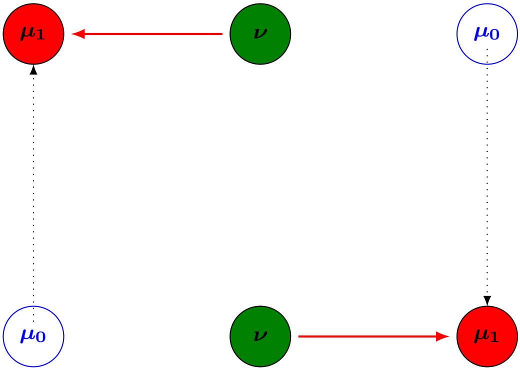
The Wasserstein space is Positively Curved (PC)

In $\mathbb{R}^m = \mathscr{P}_2(\mathbb{R}^2)$ consider two point masses $\boldsymbol{\mu_0}$ and $\boldsymbol{\mu_1}$ ... and a third reference measure $\boldsymbol{\nu}$.



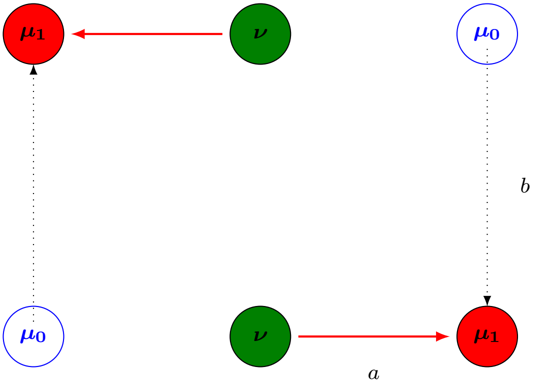
The Wasserstein space is Positively Curved (PC)

In $\mathbb{R}^m = \mathscr{P}_2(\mathbb{R}^2)$ consider two point masses $\boldsymbol{\mu_0}$ and $\boldsymbol{\mu_1}$ ... and a third reference measure $\boldsymbol{\nu}$.



The Wasserstein space is Positively Curved (PC)

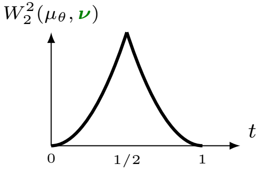
In $\mathbb{R}^m = \mathscr{P}_2(\mathbb{R}^2)$ consider two point masses $\boldsymbol{\mu_0}$ and $\boldsymbol{\mu_1}$ ...and a third reference measure $\boldsymbol{\nu}$.



The Wasserstein distance is given by

$$W_2^2(\boldsymbol{\nu}, \mu_\theta) = \min\Big(a^2 + b^2\theta^2, a^2 + b^2(1-\theta)^2\Big)$$

It is not $\boldsymbol{\lambda}$-convex, for any $\boldsymbol{\lambda}$.



Slope estimate (1)

Rewrite the minimum problem

$$\frac{W_2^2(M_\tau^n, M_\tau^{n-1})}{2\tau} + \Phi(M_\tau^n) \le \frac{W_2^2(V, M_\tau^{n-1})}{2\tau} + \Phi(V)$$

as

$$\begin{aligned}
\Phi(M_\tau^n) - \Phi(V) &\le \frac{W_2^2(V, M_\tau^{n-1})}{2\tau} - \frac{W_2^2(M_\tau^n, M_\tau^{n-1})}{2\tau} \\
&= \frac{W_2(V, M_\tau^{n-1}) - W_2(M_\tau^n, M_\tau^{n-1})}{\tau} \, \frac{W_2(V, M_\tau^{n-1}) + W_2(M_\tau^n, M_\tau^{n-}}{2} \\
(\textit{triangular inequality}) &\le \frac{W_2(V, M_\tau^n)}{\tau} \, \frac{W_2(V, M_\tau^{n-1}) + W_2(M_\tau^n, M_\tau^{n-1})}{2}.
\end{aligned}$$

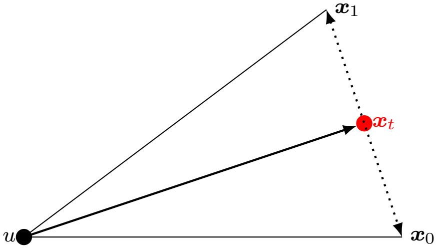
Dividing by $W_2(V, M_\tau^n)$ and passing to the limit as $V \to M_\tau^n$ we get

$$|\partial\Phi|(M_\tau^n) \le \frac{W_2(M_\tau^n, M_\tau^{n-1})}{\tau}$$

Positively Curved (PC) spaces and semiconcavity of the distance

The euclidean case $\downdownarrows$ identity

$$\mathsf{d}^2(u, \boldsymbol{x_t}) = (1-\boldsymbol{t})\mathsf{d}^2(u, \boldsymbol{x}_0) + \boldsymbol{t}\mathsf{d}^2(u, \boldsymbol{x}_1) \qquad - \ \boldsymbol{t(1-t)}\mathsf{d}^2(\boldsymbol{x}_0, \boldsymbol{x}_1).$$



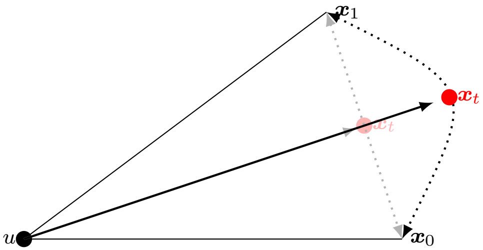
Positively Curved (PC) spaces and semiconcavity of the distance

The euclidean case $\downdownarrows$ identity

$$\mathsf{d}^2(u,\boldsymbol{x_t}) = (1-\boldsymbol{t})\mathsf{d}^2(u,\boldsymbol{x}_0) + \boldsymbol{t}\mathsf{d}^2(u,\boldsymbol{x}_1) \qquad - \ \boldsymbol{t(1-t)}\mathsf{d}^2(\boldsymbol{x}_0,\boldsymbol{x}_1).$$

The Wasserstein case $\downdownarrows$ inequality

$$W_2^2(\nu,\boldsymbol{\mu_t}) \boxed{\boxed{\ge}} (1-\boldsymbol{t})W_2^2(\nu,\mu_0) + \boldsymbol{t}W_2^2(\nu,\mu_1) \qquad - \ \boldsymbol{t(1-t)}W_2^2(\mu_0,\mu_1).$$



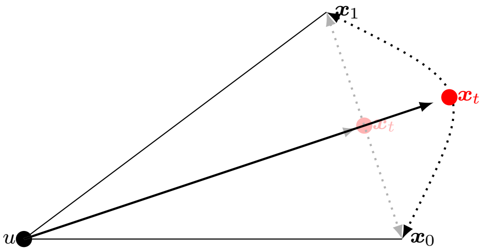
Positively Curved (PC) spaces and semiconcavity of the distance

The euclidean case $\Downarrow$ identity

$$\mathsf{d}^2(u, \boldsymbol{x_t}) = (1-\boldsymbol{t})\mathsf{d}^2(u, \boldsymbol{x}_0) + \boldsymbol{t}\mathsf{d}^2(u, \boldsymbol{x}_1) \qquad - \ \boldsymbol{t(1-t)}\mathsf{d}^2(\boldsymbol{x}_0, \boldsymbol{x}_1).$$

The Wasserstein case $\Downarrow$ inequality

$$W_2^2(\nu, \boldsymbol{\mu_t}) \boxed{\geq} (1-\boldsymbol{t})W_2^2(\nu, \mu_0) + \boldsymbol{t}W_2^2(\nu, \mu_1) \qquad - \ \boldsymbol{t(1-t)}W_2^2(\mu_0, \mu_1).$$

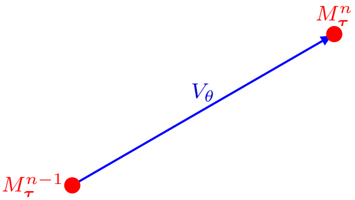


Difficulty: the functional $\quad \boldsymbol{\mu} \mapsto \dfrac{1}{2\boldsymbol{\tau}}W_2^2(\nu, \boldsymbol{\mu}) + \Phi(\boldsymbol{\mu})$

looses any convexity property

Refined discrete energy estimate (2): convex functionals

$M_{\boldsymbol{\tau}}^n$ minimizes the functional $U \mapsto \dfrac{W_2^2(U, M_{\boldsymbol{\tau}}^{n-1})}{2\boldsymbol{\tau}} + \phi(U)$



$$V_\theta := (1-\theta)M_\tau^{n-1} + \theta\, M_\tau^n \xrightarrow{\theta\uparrow 1} M_\tau^n$$

$$\frac{W_2^2(V_\theta, M_{\boldsymbol{\tau}}^{n-1})}{2\boldsymbol{\tau}} = \theta^2\frac{W_2^2(M_\tau^n, M_\tau^{n-1})}{2\tau}, \qquad \Phi(V_\theta) \le (1-\theta)\,\Phi(M_\tau^{n-1}) + \theta\,\Phi(M_\tau^n)$$

The function

$$\theta \mapsto \theta^2\frac{W_2^2(M_\tau^n, M_\tau^{n-1})}{2\tau} + (1-\theta)\,\Phi(M_\tau^{n-1}) + \theta\,\Phi(M_\tau^n)$$

has a minimum at $\theta = 1$; its derivative at $\theta = 1$ is therefore nonpositive and we find

$$\boxed{\mathbf{1}}\frac{W_2^2(M_\tau^n, M_\tau^{n-1})}{\tau} + \Phi(M_\tau^n) - \Phi(M_\tau^{n-1}) \le 0$$

Discrete slope-energy inequality, convex functionals

Slope estimate (1) $$\boxed{|\partial\Phi|(M_\tau^n) \le \frac{W_2(M_\tau^n, M_\tau^{n-1})}{\tau}}$$

Discrete energy estimate (2) $$\boxed{\frac{W_2^2(M_\tau^n, M_\tau^{n-1})}{\tau} \le \Phi(M_\tau^{n-1}) - \Phi(M_\tau^n)}$$

$$\frac{\tau}{2}\frac{W_2^2(M_\tau^n, M_\tau^{n-1})}{\tau^2} + \frac{\tau}{2}|\partial\Phi|^2(M_\tau^n) \le \Phi(M_\tau^{n-1}) - \Phi(M_\tau^n)$$

Summing up:

$$\frac{\tau}{2}\sum_{n=1}^{N}\frac{W_2^2(M_\tau^n, M_\tau^{n-1})}{\tau^2} + \frac{\tau}{2}\sum_{n=1}^{N}|\partial\Phi|^2(M_\tau^n) \le \Phi(\boldsymbol{u}_0) - \Phi(M_\tau^N)$$

Compare with the continuous energy identity

$$\frac{1}{2}\int_0^T |\dot{\boldsymbol{u}}_t|^2\,\mathrm{d}t + \frac{1}{2}\int_0^T |\partial\Phi|^2(\boldsymbol{u}_t)\,\mathrm{d}t = \Phi(\boldsymbol{u}_0) - \Phi(\boldsymbol{u}_t).$$

Slope decay (2), convex functionals

$$\boxed{\frac{\tau}{2}|\partial\Phi|^2(M_\tau^n) + \frac{\tau}{2}\frac{W_2^2(M_\tau^n, M_\tau^{n-1})}{\tau^2} \le \Phi(M_\tau^{n-1}) - \Phi(M_\tau^n)}$$

By the definition of the slope for a convex function

$$\Phi(M_\tau^{n-1}) - \Phi(M_\tau^n) \le |\partial\Phi|(M_\tau^{n-1})\, W_2(M_\tau^{n-1}, M_\tau^n)$$

$$\begin{aligned}\frac{\tau}{2}|\partial\Phi|^2(M_\tau^n) + \frac{\tau}{2}\frac{W_2^2(M_\tau^n, M_\tau^{n-1})}{\tau^2} &\le \Phi(M_\tau^{n-1}) - \Phi(M_\tau^n) \le |\partial\Phi|(M_\tau^{n-1})\, W_2(M_\tau^{n-1}, M_\tau^n) \\ &\le \frac{\tau}{2}|\partial\Phi|^2(M_\tau^{n-1}) + \frac{1}{2\tau}W_2^2(M_\tau^{n-1}, M_\tau^n)\end{aligned}$$

$$\boxed{|\partial\Phi|(M_\tau^n) \le |\partial\Phi|(M_\tau^{n-1})}$$

Summary

- **Rough discrete energy inequality**

$$\frac{\tau}{2}\frac{W_2^2(M_\tau^n, M_\tau^{n-1})}{\tau^2} \le \Phi(M_\tau^{n-1}) - \Phi(M_\tau^n)$$

- **Slope estimate**

$$|\partial\Phi|(M_\tau^n) \le \frac{W_2(M_\tau^n, M_\tau^{n-1})}{\tau}$$

Convexity:

- **Discrete slope–energy inequality**

$$\frac{\tau}{2}\frac{W_2^2(M_\tau^n, M_\tau^{n-1})}{\tau^2} + \frac{\tau}{2}|\partial\Phi|^2(M_\tau^n) \le \Phi(M_\tau^{n-1}) - \Phi(M_\tau^n)$$

- **Slope decay**

$$|\partial\Phi|(M_\tau^n) \le |\partial\Phi|(M_\tau^{n-1})$$

Convergence for convex functionals with compact sublevels

If $\Phi : \mathscr{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ be a convex functional with compact sublevels then the piecewise constant interpolant $\boldsymbol{M_\tau}$ converge pointwise in $\mathscr{P}_2(\mathbb{R}^d)$. Passing to the limit by lower semicontinuity in the inequality

$$\frac{\tau}{2}\sum_{n=1}^{N}\frac{W_2^2(M_\tau^n, M_\tau^{n-1})}{\tau^2} + |\partial\Phi|^2(M_\tau^n) + \Phi(M_\tau^N) \leq \Phi(M_\tau^0)$$

we get

$$\frac{1}{2}\int_0^T |\dot\mu_t|^2 + |\partial\Phi|^2(\mu_t)\,\mathrm{d}t + \Phi(\mu_T) \leq \Phi(\mu_0)$$

which is enough to conclude.