UniGe | MaLGa

Inexact proximal-gradient algorithm in the Wasserstein space: links and differences from the Hilbert case

Emanuele Naldi
(Joint work with Simone Di Marino and Silvia Villa)

MaLGa- Machine learning Genova Center
University of Genoa

18 September 2025

Outline

Inexact JKO

Inexact proximal-gradient

Towards nonexpansivity of the proximal map

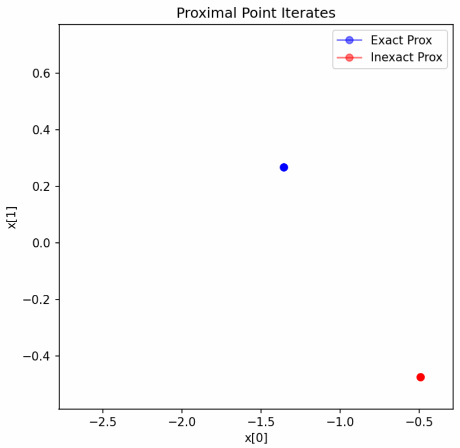
Inexact JKO

Given $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ consider

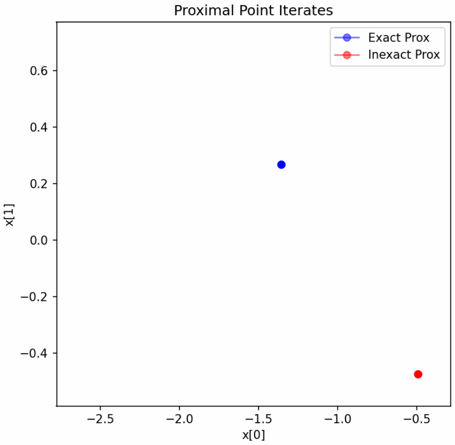
$$J_\tau(\mu) = \underset{\nu \in \mathcal{P}_2(\mathbb{R}^d)}{\arg\min}\, \mathcal{G}(\nu) + \frac{1}{2\tau} W_2^2(\nu, \mu), \tag{1}$$

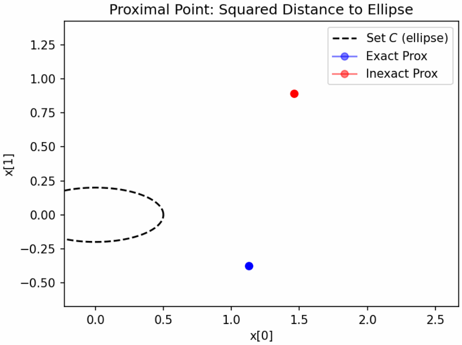
and the algorithm defined by $\mu_{n+1} \approx J_\tau(\mu_n)$.

Inexact proximal point algorithm

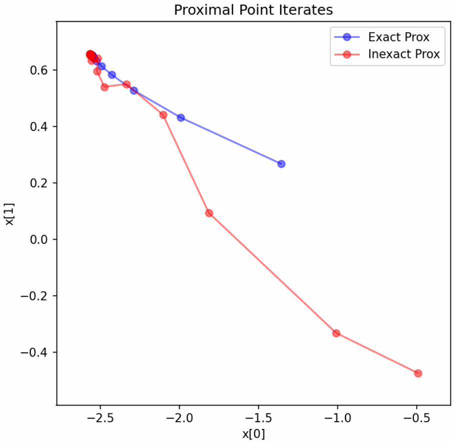


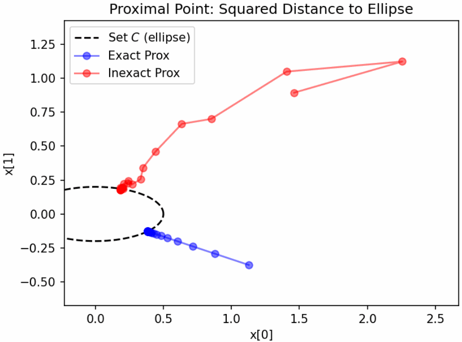
Inexact proximal point algorithm





UniGe | MaLGa


Proximal Point Iterates
Exact Prox
Inexact Prox
0.6
0.4
0.2
0.0
−0.2
−0.4
−2.5
−2.0
−1.5
−1.0
−0.5
x[0]
x[1]


Proximal Point: Squared Distance to Ellipse
Set C (ellipse)
Exact Prox
Inexact Prox
1.25
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
0.0
0.5
1.0
1.5
2.0
2.5
x[0]
x[1]

JKO

Given $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ consider

$$J_\tau(\mu) = \underset{\nu \in \mathcal{P}_2(\mathbb{R}^d)}{\arg\min}\, \mathcal{G}(\nu) + \frac{1}{2\tau} W_2^2(\nu, \mu), \tag{1}$$

and the algorithm defined by $\mu_{n+1} = J_\tau(\mu_n)$.

JKO

Given $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ consider

$$J_\tau(\mu) = \operatorname*{arg\,min}_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \mathcal{G}(\nu) + \frac{1}{2\tau} W_2^2(\nu, \mu), \tag{1}$$

and the algorithm defined by $\mu_{n+1} = J_\tau(\mu_n)$.

Theorem (N., Savaré '22)

Let $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ proper, lower semicontinuous, convex along generalized geodesics and $\arg\min \mathcal{G} \neq \emptyset$. Then the sequence $\mu_n \rightharpoonup \mu^* \in \arg\min \mathcal{G}$ narrowly.

JKO

Given $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ consider

$$J_\tau(\mu) = \operatorname*{arg\,min}_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \mathcal{G}(\nu) + \frac{1}{2\tau} W_2^2(\nu, \mu), \tag{1}$$

and the algorithm defined by $\mu_{n+1} = J_\tau(\mu_n)$.

Theorem (N., Savaré '22)

Let $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ proper, lower semicontinuous, convex along generalized geodesics and $\operatorname{arg\,min} \mathcal{G} \neq \emptyset$. Then $\mu_n \overset{w,2}{\rightharpoonup} \mu^* \in \operatorname{arg\,min} \mathcal{G}$.

JKO

Given $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ consider

$$J_\tau(\mu) = \underset{\nu \in \mathcal{P}_2(\mathbb{R}^d)}{\arg\min}\, \mathcal{G}(\nu) + \frac{1}{2\tau} W_2^2(\nu, \mu), \tag{1}$$

and the algorithm defined by $\mu_{n+1} = J_\tau(\mu_n)$.

Theorem (N., Savaré '22)

Let $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ proper, lower semicontinuous, convex along generalized geodesics and $\arg\min \mathcal{G} \neq \emptyset$. Then $\mu_n \xrightarrow{w,2} \mu^* \in \arg\min \mathcal{G}$.

To prove this we used

- Opial property in Wasserstein spaces [N. and Savaré, 2022]
- Every lower semicontinuous and geodesically convex functional is sequentially lower semicontinuous w.r.t. the topology $\tau_{w,2}$ [N. and Savaré, 2022]

Topological setting

Let $C_2^w(\mathbb{R}^d)$ be the space defined by

$$C_2^w(\mathbb{R}^d) := \left\{ f : \mathbb{R}^d \to \mathbb{R} \;\middle|\; f \text{ is continuous and } \lim_{\|x\|\to\infty} \frac{f(x)}{1+\|x\|^2} = 0 \right\},$$

endowed with the norm $\|f\|_{C_2^w(\mathbb{R}^d)} := \sup_{x\in\mathbb{R}^d} \frac{|f(x)|}{1+\|x\|^2}$. The space

$$\mathcal{M}_2(\mathbb{R}^d) = \left\{ \mu \in \mathcal{M}(\mathbb{R}^d) \;\middle|\; \int_{\mathbb{R}^d} \|x\|^2 \, d|\mu|(x) < +\infty \right\},$$

with the norm $\|\mu\|_{\mathcal{M}_2(\mathbb{R}^d)} = \int_{\mathbb{R}^d} 1 + \|x\|^2 \, d|\mu|(x)$ is the dual of $C_2^w(\mathbb{R}^d)$.

Topological setting

Let $C_2^w(\mathbb{R}^d)$ be the space defined by

$$C_2^w(\mathbb{R}^d) := \left\{ f : \mathbb{R}^d \to \mathbb{R} \;\middle|\; f \text{ is continuous and } \lim_{\|x\|\to\infty} \frac{f(x)}{1+\|x\|^2} = 0 \right\},$$

endowed with the norm $\|f\|_{C_2^w(\mathbb{R}^d)} := \sup_{x\in\mathbb{R}^d} \frac{|f(x)|}{1+\|x\|^2}$. The space

$$\mathcal{M}_2(\mathbb{R}^d) = \left\{ \mu \in \mathcal{M}(\mathbb{R}^d) \;\middle|\; \int_{\mathbb{R}^d} \|x\|^2 \, d|\mu|(x) < +\infty \right\},$$

with the norm $\|\mu\|_{\mathcal{M}_2(\mathbb{R}^d)} = \int_{\mathbb{R}^d} 1 + \|x\|^2 \, d|\mu|(x)$ is the dual of $C_2^w(\mathbb{R}^d)$.

We consider the weak-$*$ topology on this space restricted to $\mathcal{P}_2(\mathbb{R}^d)$. We denote the convergence by $\mu_n \xrightarrow{w,2} \mu$.

- The topology is finer than the narrow topology
- It implies convergence in p-Wasserstein distance for any $p \in [1, 2)$.

Definition (Generalized geodesic)

A generalized geodesic between $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ (with base $\nu \in \mathcal{P}_2(\mathbb{R}^d)$) is a curve $(\mu_t)_{t\in[0,1]}$ in $\mathcal{P}_2(\mathbb{R}^d)$ defined by

$$\mu_t = (\pi_t^{2\to 3})_{\#}\gamma \quad t \in [0,1],$$

where $\pi_t^{2\to 3} := (1-t)\pi^2 + t\pi^3$, $\gamma \in \Gamma(\nu, \mu_0, \mu_1)$, $\pi^{1,2}_{\#}\gamma \in \Gamma_{\mathrm{opt}}(\nu, \mu_0)$ and $\pi^{1,3}_{\#}\gamma \in \Gamma_{\mathrm{opt}}(\nu, \mu_1)$.

Definition (Convexity)

$\mathcal{G}$ is convex along (generalized) geodesics if for every $\mu_0, \mu_1, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ there exists a (generalized) geodesic $(\mu_s)_{s\in[0,1]}$ between μ_0 and μ_1 (with base ν), along which

$$\mathcal{G}(\mu_s) \leqslant (1-s)\mathcal{G}(\mu_0) + s\mathcal{G}(\mu_1) \quad \text{for every } s \in [0,1].$$

Definition (Generalized geodesic)

A generalized geodesic between $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ (with base $\nu \in \mathcal{P}_2(\mathbb{R}^d)$) is a curve $(\mu_t)_{t\in[0,1]}$ in $\mathcal{P}_2(\mathbb{R}^d)$ defined by

$$\mu_t = (\pi_t^{2\to 3})_{\#}\gamma \quad t \in [0,1],$$

where $\pi_t^{2\to 3} := (1-t)\pi^2 + t\pi^3$, $\gamma \in \Gamma(\nu, \mu_0, \mu_1)$, $\pi_{\#}^{1,2}\gamma \in \Gamma_{\mathsf{opt}}(\nu, \mu_0)$ and $\pi_{\#}^{1,3}\gamma \in \Gamma_{\mathsf{opt}}(\nu, \mu_1)$.

Definition (Convexity)

$\mathcal{G}$ is convex along (generalized) geodesics if for every $\mu_0, \mu_1, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ there exists a (generalized) geodesic $(\mu_s)_{s\in[0,1]}$ between μ_0 and μ_1 (with base ν), along which

$$\mathcal{G}(\mu_s) \leqslant (1-s)\mathcal{G}(\mu_0) + s\mathcal{G}(\mu_1) \quad \text{for every } s \in [0,1].$$

Definition (Generalized geodesic)

A generalized geodesic between $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ (with base $\nu \in \mathcal{P}_2(\mathbb{R}^d)$) is a curve $(\mu_t)_{t\in[0,1]}$ in $\mathcal{P}_2(\mathbb{R}^d)$ defined by

$$\mu_t = (\pi_t^{2\to 3})_{\#}\gamma \quad t \in [0,1],$$

where $\pi_t^{2\to 3} := (1-t)\pi^2 + t\pi^3$, $\gamma \in \Gamma(\nu, \mu_0, \mu_1)$, $\pi^{1,2}_{\#}\gamma \in \Gamma_{\mathsf{opt}}(\nu, \mu_0)$ and $\pi^{1,3}_{\#}\gamma \in \Gamma_{\mathsf{opt}}(\nu, \mu_1)$.

Definition (Convexity)

$\mathcal{G}$ is convex along (generalized) geodesics if for every $\mu_0, \mu_1, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ there exists a (generalized) geodesic $(\mu_s)_{s\in[0,1]}$ between μ_0 and μ_1 (with base ν), along which

$$\mathcal{G}(\mu_s) \leqslant (1-s)\mathcal{G}(\mu_0) + s\mathcal{G}(\mu_1) \quad \text{for every } s \in [0,1].$$

Definition (Generalized geodesic)

A generalized geodesic between $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ (with base $\nu \in \mathcal{P}_2(\mathbb{R}^d)$) is a curve $(\mu_t)_{t\in[0,1]}$ in $\mathcal{P}_2(\mathbb{R}^d)$ defined by

$$\mu_t = (\pi_t^{2\to 3})_{\#}\gamma \quad t \in [0,1],$$

where $\pi_t^{2\to 3} := (1-t)\pi^2 + t\pi^3$, $\gamma \in \Gamma(\nu, \mu_0, \mu_1)$, $\pi^{1,2}_{\#}\gamma \in \Gamma_{\mathsf{opt}}(\nu, \mu_0)$ and $\pi^{1,3}_{\#}\gamma \in \Gamma_{\mathsf{opt}}(\nu, \mu_1)$.

Definition (Convexity)

$\mathcal{G}$ is convex along (generalized) geodesics if for every $\mu_0, \mu_1, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ there exists a (generalized) geodesic $(\mu_s)_{s\in[0,1]}$ between μ_0 and μ_1 (with base ν), along which

$$\mathcal{G}(\mu_s) \leqslant (1-s)\mathcal{G}(\mu_0) + s\mathcal{G}(\mu_1) \quad \text{for every } s \in [0,1].$$

Theorem

Let $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous and convex along generalized geodesics with $\arg\min \mathcal{G} \neq \emptyset$. Then, for each $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, it holds

$$W_2^2(J_\tau(\mu), \nu) - W_2^2(\mu, \nu) \leqslant 2\tau\left(\mathcal{G}(\nu) - \mathcal{G}(J_\tau(\mu))\right) - W_2^2(J_\tau(\mu), \mu)$$

Theorem

Let $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous and convex along generalized geodesics with $\arg\min \mathcal{G} \neq \emptyset$. Then, for each $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, it holds

$$W_2^2(J_\tau(\mu), \nu) - W_2^2(\mu, \nu) \leqslant 2\tau \left(\mathcal{G}(\nu) - \mathcal{G}(J_\tau(\mu))\right) - W_2^2(J_\tau(\mu), \mu)$$

Theorem

Let $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ proper, lower semicontinuous and convex along generalized geodesics with $\arg\min \mathcal{G} \neq \emptyset$. Then, for each $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, it holds

$$W_2^2(J_\tau(\mu), \nu) - W_2^2(\mu, \nu) \leqslant 2\tau \left(\mathcal{G}(\nu) - \mathcal{G}(J_\tau(\mu))\right) - W_2^2(J_\tau(\mu), \mu)$$

Theorem (Di Marino, N., Villa)

Let $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ proper, lower semicontinuous and convex along generalized geodesics with $\arg\min \mathcal{G} \neq \emptyset$. Let $\{\mu_n\}_n$ satisfying

$$W_2(\mu_{n+1}, J_\tau(\mu_n)) \leqslant \epsilon_n, \quad \text{for all } n \in \mathbb{N}, \tag{2}$$

with $\sum_n \epsilon_n < +\infty$ and ϵ_n not increasing. Then

$$\mathcal{G}(J_\tau(\mu_n)) - \inf \mathcal{G} = O\left(\frac{1}{n}\right), \quad \text{as } n \to \infty,$$

and $\mu_n \xrightarrow{w,2} \mu^* \in \arg\min \mathcal{G}$.

Theorem (Di Marino, N., Villa)

Let $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ be proper, lower semicontinuous, and convex along generalized geodesics, with $\arg\min \mathcal{G} \neq \emptyset$. Let $\{\epsilon_n\}_n \subset \mathbb{R}_{\geqslant 0}$ with $\sum_{n=0}^{\infty} \epsilon_n < \infty$ and let $\{\tau_n\}_n \subset \mathbb{R}_{>0}$ with $\sum_{i=0}^{\infty} \tau_i = \infty$. Define $\sigma_n := \sum_{i=0}^{n-1} \tau_i$, for $n \in \mathbb{N}$. Let $\{\mu_n\}_n$ be a sequence satisfying

$$W_2(\mu_{n+1}, J_{\tau_n}(\mu_n)) \leqslant \epsilon_n, \quad \text{for all } n \in \mathbb{N}.$$

1. It holds the rate

$$\mathcal{G}(\bar{\beta}_n) - \inf \mathcal{G} = O\left(\frac{1}{\sigma_n}\right), \quad \text{as } n \to \infty,$$

   where $\bar{\beta}_n := J_{\tau_{j_n}}(\mu_{j_n})$ with $j_n = \arg\min_{i=0,\dots,n-1}\{\mathcal{G}(J_{\tau_i}(\mu_i))\}$, defines the sequence of the best iterates.
2. If $\sum_{n=1}^{\infty} \frac{\sigma_n}{\tau_n} \epsilon_{n-1}^2 < \infty$, then

$$\mathcal{G}(J_{\tau_n}(\mu_n)) - \inf \mathcal{G} = O\left(\frac{1}{\sigma_n}\right), \quad \text{as } n \to \infty,$$

   and $\mu_n \xrightarrow{w,2} \mu^* \in \arg\min \mathcal{G}$.

Theorem (Di Marino, N., Villa)

Let $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ proper, lower semicontinuous and convex along generalized geodesics with $\arg\min \mathcal{G} \neq \emptyset$. Let $\{\mu_n\}_n$ satisfying

$$\mathcal{G}(\mu_{n+1}) + \frac{1}{2\tau} W_2^2(\mu_{n+1}, \mu_n) \leqslant \mathcal{G}(J_\tau(\mu_n)) + \frac{1}{2\tau} W_2^2(J_\tau(\mu_n), \mu_n) + \epsilon_n^2. \tag{3}$$

with $\sum_n \epsilon_n < +\infty$ and ϵ_k not increasing. Then

$$\mathcal{G}(\mu_n) - \inf \mathcal{G} = O\left(\frac{1}{n}\right), \quad \text{as } n \to \infty,$$

and $\mu_n \xrightarrow{w,2} \mu^* \in \arg\min \mathcal{G}$.

Theorem (Di Marino, N., Villa)

Let $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ proper, lower semicontinuous and convex along generalized geodesics with $\arg\min \mathcal{G} \neq \emptyset$. Let $\{\mu_n\}_n$ satisfying

$$\mathcal{G}(\mu_{n+1}) + \frac{1}{2\tau} W_2^2(\mu_{n+1}, \mu_n) \leqslant \mathcal{G}(J_\tau(\mu_n)) + \frac{1}{2\tau} W_2^2(J_\tau(\mu_n), \mu_n) + \epsilon_n^2. \tag{3}$$

with $\sum_n \epsilon_n < +\infty$ and ϵ_k not increasing. Then

$$\mathcal{G}(\mu_n) - \inf \mathcal{G} = O\left(\frac{1}{n}\right), \quad \text{as } n \to \infty,$$

and $\mu_n \xrightarrow{w,2} \mu^* \in \arg\min \mathcal{G}$.

- Condition (3) implies also (2).

Theorem (Di Marino, N., Villa)

Let $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ proper, lower semicontinuous and convex along generalized geodesics with $\arg\min \mathcal{G} \neq \emptyset$. Let $\{\mu_n\}_n$ satisfying

$$\mathcal{G}(\mu_{n+1}) + \frac{1}{2\tau} W_2^2(\mu_{n+1}, \mu_n) \leqslant \mathcal{G}(J_\tau(\mu_n)) + \frac{1}{2\tau} W_2^2(J_\tau(\mu_n), \mu_n) + \epsilon_n^2. \tag{3}$$

with $\sum_n \epsilon_n < +\infty$ and ϵ_k not increasing. Then

$$\mathcal{G}(\mu_n) - \inf \mathcal{G} = O\left(\frac{1}{n}\right), \quad \text{as } n \to \infty,$$

and $\mu_n \xrightarrow{w,2} \mu^* \in \arg\min \mathcal{G}$.

- Condition (3) implies also (2).
- Here, the proof is more straightforward than the previous one.

Theorem (Di Marino, N., Villa)

Let $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ proper, lower semicontinuous and convex along generalized geodesics with $\arg\min \mathcal{G} \neq \varnothing$. Let $\{\mu_n\}_n$ satisfying

$$\mathcal{G}(\mu_{n+1}) + \frac{1}{2\tau} W_2^2(\mu_{n+1}, \mu_n) \leqslant \mathcal{G}(J_\tau(\mu_n)) + \frac{1}{2\tau} W_2^2(J_\tau(\mu_n), \mu_n) + \epsilon_n^2. \tag{3}$$

with $\sum_n \epsilon_n < +\infty$ and ϵ_k not increasing. Then

$$\mathcal{G}(\mu_n) - \inf \mathcal{G} = O\left(\frac{1}{n}\right), \quad \text{as } n \to \infty,$$

and $\mu_n \xrightarrow{w,2} \mu^* \in \arg\min \mathcal{G}$.

- Condition (3) implies also (2).
- Here, the proof is more straightforward than the previous one.
- We do not need further assumptions on $\mathcal{G}$ to get rates on $\{\mathcal{G}(\mu_n)\}_n$.

Outline

Inexact JKO

Inexact proximal-gradient

Towards nonexpansivity of the proximal map

Extension to Proximal-Gradient

Consider the problem

$$\min_{\mu\in\mathcal{P}_2(X)} \mathcal{G}(\mu) = \int F\, d\mu + \mathcal{H}(\mu),$$

(A1) $F\colon \mathbb{R}^d \to \mathbb{R}$ is convex and differentiable with L-Lipschitz continuous gradient and convex

(A2) $\mathcal{H}\colon \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}\cup\{+\infty\}$ is proper, lower semicontinuous and convex along generalized geodesics

(A3) $\mathrm{dom}(\mathcal{H}) \subset \mathcal{P}_2^r(\mathcal{X})$.

Extension to Proximal-Gradient

Consider the problem

$$\min_{\mu \in \mathcal{P}_2(X)} \mathcal{G}(\mu) = \int F \, d\mu + \mathcal{H}(\mu),$$

(A1) $F \colon \mathbb{R}^d \to \mathbb{R}$ is convex and differentiable with L-Lipschitz continuous gradient and convex

(A2) $\mathcal{H} \colon \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ is proper, lower semicontinuous and convex along generalized geodesics

(A3) $\mathrm{dom}(\mathcal{H}) \subset \mathcal{P}_2^r(\mathcal{X})$.

Define the operator $\mathcal{T} := J_{\tau\mathcal{H}} \circ (I - \tau \nabla F)_{\#}$ and consider a scheme that satisfy

$$W_2(\mu_{n+1}, \mathcal{T}(\mu_n)) \leqslant \epsilon_n, \quad \text{for all } n \in \mathbb{N}. \tag{4}$$

Extension to Proximal-Gradient

Consider the problem

$$\min_{\mu\in\mathcal{P}_2(X)} \mathcal{G}(\mu) = \int F\, d\mu + \mathcal{H}(\mu),$$

(A1) $F\colon \mathbb{R}^d \to \mathbb{R}$ is convex and differentiable with L-Lipschitz continuous gradient and convex

(A2) $\mathcal{H}\colon \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}\cup\{+\infty\}$ is proper, lower semicontinuous and convex along generalized geodesics

(A3) $\mathrm{dom}(\mathcal{H}) \subset \mathcal{P}_2^r(\mathcal{X})$.

Define the operator $\mathcal{T} := J_{\tau\mathcal{H}} \circ (I - \tau\nabla F)_{\#}$ and consider a scheme that satisfy

$$W_2(\mu_{n+1}, \mathcal{T}(\mu_n)) \leqslant \epsilon_n, \quad \text{for all } n \in \mathbb{N}. \tag{4}$$

We know by [Salim, Korba and Luise, 2020] that for all $\mu, \nu \in \mathcal{P}_2^r(\mathbb{R}^d)$ it holds

$$W_2^2(\mathcal{T}(\mu), \nu) \leqslant W_2^2(\mu, \nu) - 2\tau(\mathcal{G}(\mathcal{T}(\mu)) - \mathcal{G}(\nu)) \tag{5}$$

Extension to Proximal-Gradient

Consider the problem

$$\min_{\mu \in \mathcal{P}_2(X)} \mathcal{G}(\mu) = \int F \, d\mu + \mathcal{H}(\mu),$$

(A1) $F \colon \mathbb{R}^d \to \mathbb{R}$ is convex and differentiable with L-Lipschitz continuous gradient and convex

(A2) $\mathcal{H} \colon \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ is proper, lower semicontinuous and convex along generalized geodesics

(A3) $\mathrm{dom}(\mathcal{H}) \subset \mathcal{P}_2^r(\mathcal{X})$.

Define the operator $\mathcal{T} := J_{\tau\mathcal{H}} \circ (I - \tau \nabla F)_\#$ and consider a scheme that satisfy

$$W_2(\mu_{n+1}, \mathcal{T}(\mu_n)) \leqslant \epsilon_n, \quad \text{for all } n \in \mathbb{N}. \tag{4}$$

Actually, **it is possible to prove** that for all $\mu, \nu \in \mathcal{P}_2^r(\mathbb{R}^d)$ it holds

$$W_2^2(\mathcal{T}(\mu), \nu) \leqslant W_2^2(\mu, \nu) - 2\tau(\mathcal{G}(\mathcal{T}(\mu)) - \mathcal{G}(\nu)) - (1 - \gamma L) W_2^2(\mu, \mathcal{T}(\mu)).$$

Extension to Proximal-Gradient

Consider the problem

$$\min_{\mu\in\mathcal{P}_2(X)} \mathcal{G}(\mu) = \int F\, d\mu + \mathcal{H}(\mu),$$

(A1) $F\colon \mathbb{R}^d \to \mathbb{R}$ is convex and differentiable with L-Lipschitz continuous gradient and convex

(A2) $\mathcal{H}\colon \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ is proper, lower semicontinuous and convex along generalized geodesics

(A3) $\mathrm{dom}(\mathcal{H}) \subset \mathcal{P}_2^r(\mathcal{X})$.

Define the operator $\mathcal{T} := J_{\tau\mathcal{H}} \circ (I - \tau\nabla F)_\#$ and consider a scheme that satisfy

$$W_2(\mu_{n+1}, \mathcal{T}(\mu_n)) \leqslant \epsilon_n, \quad \text{for all } n \in \mathbb{N}. \tag{4}$$

Actually, **it is possible to prove** that for all $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ it holds

$$W_2^2(\mathcal{T}(\mu), \nu) \leqslant W_2^2(\mu, \nu) - 2\tau(\mathcal{G}(\mathcal{T}(\mu)) - \mathcal{G}(\nu)) - (1 - \gamma L)W_2^2(\mu, \mathcal{T}(\mu)).$$

Convergence [Di Marino, N., Villa]

Theorem (Convergence with distance error)

Let $\mathcal{H} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ and F satisfying (A1)-(A3) and suppose $\arg\min \mathcal{G} \neq \emptyset$. Let $\mu^0 \in \mathcal{P}_2^r(X)$, $\tau < 1/L$ and $\{\mu_n\}_n$ satisfying

$$\mu_{n+1} \in \mathcal{P}_2^r(X), \quad W_2\left(\mu_{n+1}, J_{\tau\mathcal{H}}((I - \tau\nabla F)_\#(\mu^k))\right) \leqslant \epsilon_n, \quad \text{for all } k \in N,$$

with $\sum_n \epsilon_n < +\infty$ and ϵ_n not increasing. Then $\mathcal{G}(\mathcal{T}(\mu_n)) - \inf \mathcal{G} = O\left(\frac{1}{n}\right)$, as $n \to \infty$, and $\mu_n \xrightarrow{w,2} \mu^* \in \arg\min \mathcal{G}$.

Convergence [Di Marino, N., Villa]

Theorem (Convergence with distance error)

Let $\mathcal{H} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ and F satisfying (A1)-(A3) and suppose $\arg\min \mathcal{G} \neq \emptyset$. Let $\mu^0 \in \mathcal{P}_2^r(X)$, $\tau < 1/L$ and $\{\mu_n\}_n$ satisfying

$$\cancel{\mu_{n+1} \in \mathcal{P}_2^r(X),} \quad W_2\left(\mu_{n+1}, J_{\tau\mathcal{H}}((I - \tau\nabla F)_\#(\mu^k))\right) \leqslant \epsilon_n, \quad \text{for all } k \in N,$$

with $\sum_n \epsilon_n < +\infty$ and ϵ_n not increasing. Then $\mathcal{G}(\mathcal{T}(\mu_n)) - \inf \mathcal{G} = O\left(\frac{1}{n}\right)$, as $n \to \infty$, and $\mu_n \xrightarrow{w,2} \mu^* \in \arg\min \mathcal{G}$.

Convergence [Di Marino, N., Villa]

Theorem (Convergence with distance error)

Let $\mathcal{H} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ and F satisfying (A1)-(A3) and suppose $\arg\min \mathcal{G} \neq \emptyset$. Let $\mu^0 \in \mathcal{P}_2^r(X)$, $\tau < 1/L$ and $\{\mu_n\}_n$ satisfying

$$W_2\left(\mu_{n+1}, J_{\tau\mathcal{H}}((I - \tau\nabla F)_\#(\mu^k))\right) \leqslant \epsilon_n, \quad \text{for all } k \in N,$$

with $\sum_n \epsilon_n < +\infty$ and ϵ_n not increasing. Then $\mathcal{G}(\mathcal{T}(\mu_n)) - \inf \mathcal{G} = O\left(\frac{1}{n}\right)$, as $n \to \infty$, and $\mu_n \overset{w,2}{\to} \mu^* \in \arg\min \mathcal{G}$.

Theorem (Convergence with variational error)

Let $\{\nu^k\}_n$, $\{\mu_n\}_n$ satisfy $\nu_{n+1} = (I - \tau\nabla F)_\#(\mu^k)$ and

$$\mathcal{H}(\mu_{n+1}) + \frac{1}{2\tau}W_2^2(\mu_{n+1}, \nu_{n+1}) \leqslant \mathcal{H}(J_\tau(\nu_{n+1})) + \frac{1}{2\tau}W_2^2(J_\tau(\nu_{n+1}), \nu_{n+1}) + \epsilon_n^2.$$

with ϵ_n as above. Then $\mathcal{G}(\mu_n) - \inf \mathcal{G} = O\left(\frac{1}{n}\right)$, as $n \to \infty$, and $\mu_n \overset{w,2}{\to} \mu^* \in \arg\min \mathcal{G}$.

Outline

Inexact JKO

Inexact proximal-gradient

Towards nonexpansivity of the proximal map

Hilbert case

- **If** g proper, convex, lsc and and $\tau \in (0, +\infty)$, **then** the proximity operator is firmly-nonexpansive:

$$\| \operatorname{prox}_{\tau g}(x) - \operatorname{prox}_{\tau g}(y)\|^2 \leqslant \|x - y\|^2 - \|(I - \operatorname{prox}_{\tau g})(x) - (I - \operatorname{prox}_{\tau g})(y)\|^2.$$

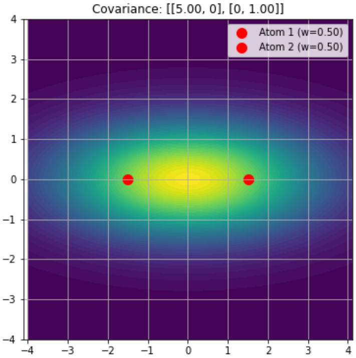
- **If** f, g proper, convex, lsc, ∇f L-Lipschitz and $\tau \in (0, 2/L)$, **then** the operator $T := \operatorname{prox}_{\tau g} \circ (I - \tau \nabla f)$ is α-averaged for some $\alpha \in (0, 1)$:

$$\|T(x) - T(y)\|^2 \leqslant \|x - y\|^2 - \frac{1-\alpha}{\alpha}\|(I - T)(x) - (I - T)(y)\|^2.$$

Wasserstein case

Location problem: given $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, we search for

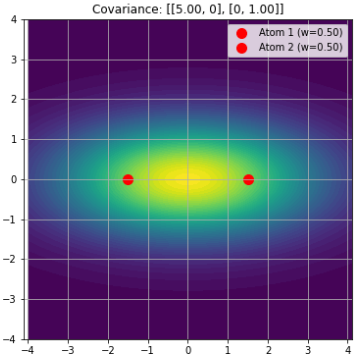
$$\arg\min \left\{ W_2(\nu, \mu) \mid \nu \in \mathcal{P}_2(\mathbb{R}^d),\ \#\operatorname{supp}(\nu) \leqslant 2 \right\}$$



Wasserstein case

Location problem: given $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, we search for

$$\arg\min \left\{ W_2(\nu, \mu) \mid \nu \in \mathcal{P}_2(\mathbb{R}^d),\ \#\operatorname{supp}(\nu) \leqslant 2 \right\}$$



In this case it is clear that $W_2(J_\tau(\mu_1), J_\tau(\mu_2)) \nleqslant W_2(\mu_1, \mu_2)$.

UniGe | MaLGa

Wasserstein case

Given the set

$$K_1 := \{\rho \in L_1^+(\Omega) \mid \int \rho(x)\,dx = 1,\ \rho \leqslant 1\},$$

and $P_{K_1}(\nu) = \arg\min\{W_2(\rho,\nu) \mid \rho \in K_1\}$.

- In [De Philippis, Mészáros, Santambrogio and Velichkov, 2016] it is proven that P_{K_1} is locally $\frac{1}{2}$-Hölder.
- In [Remark 5.1] they say that Lipschitzianity is still open.

Wasserstein case [Cavagnari, Savaré and Sodini, 2023]

Definition (λ-totally convex functionals)

Let $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ be a proper function. $\mathcal{G}$ is totally convex if for every $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ and $\gamma \in \Gamma(\mu_0, \mu_1)$ it holds

$$\mathcal{G}(\pi^t_{\#}\gamma) \leqslant (1-s)\mathcal{G}(\mu_0) + s\mathcal{G}(\mu_1) \quad \text{for every } s \in [0,1].$$

Wasserstein case [Cavagnari, Savaré and Sodini, 2023]

Definition (λ-totally convex functionals)

Let $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ be a proper function. $\mathcal{G}$ is totally convex if for every $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ and $\gamma \in \Gamma(\mu_0, \mu_1)$ it holds

$$\mathcal{G}(\pi^t_{\#}\gamma) \leqslant (1-s)\mathcal{G}(\mu_0) + s\mathcal{G}(\mu_1) \quad \text{for every } s \in [0,1].$$

Theorem

Let $\mathcal{G}$ proper, lsc and totally convex, one can show that it holds

$$W_2(J_\tau(\mu), J_\tau(\nu)) \leqslant W_2(\mu, \nu), \quad \text{for all } \mu, \nu \in \mathcal{P}_2(X).$$

Wasserstein case [Cavagnari, Savaré and Sodini, 2023]

Definition (λ-totally convex functionals)

Let $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ be a proper function. $\mathcal{G}$ is totally convex if for every $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ and $\gamma \in \Gamma(\mu_0, \mu_1)$ it holds

$$\mathcal{G}(\pi^t_{\#}\gamma) \leqslant (1-s)\mathcal{G}(\mu_0) + s\mathcal{G}(\mu_1) \quad \text{for every } s \in [0,1].$$

Theorem

Let $\mathcal{G}$ proper, lsc and totally convex, one can show that it holds

$$W_2(J_\tau(\mu), J_\tau(\nu)) \leqslant W_2(\mu, \nu), \quad \text{for all } \mu, \nu \in \mathcal{P}_2(X).$$

Examples:

- Potential energy and internal energy are totally convex
- The indicator function of $\{\nu \in \mathcal{P}_2(\mathbb{R}^d) \mid \#\operatorname{supp}(\nu) \leqslant 2\}$ is not totally convex
- The indicator function of K_1 is not totally convex

How many are they?

Theorem (Many! Cavagnari, Savaré, Sodini)

Assume that $d \geqslant 2$, $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ is a proper l.s.c. geodesically convex functional such that the discrete measures are dense in energy, i.e., for every $\mu \in \mathrm{dom}(\mathcal{G})$ there exists a sequence $\{\mu_n\}_n$ of discrete measures such that

$$\mu_n \to \mu \quad \text{and} \quad \mathcal{G}(\mu_n) \to \mathcal{G}(\mu).$$

Then $\mathcal{G}$ is totally convex.

UniGe | MaLGa

How many are they?

Theorem (Many! Cavagnari, Savaré, Sodini)

Assume that $d \geqslant 2$, $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ is a proper l.s.c. geodesically convex functional such that the discrete measures are dense in energy, i.e., for every $\mu \in$ dom$(\mathcal{G})$ there exists a sequence $\{\mu_n\}_n$ of discrete measures such that

$$\mu_n \to \mu \quad \text{and} \quad \mathcal{G}(\mu_n) \to \mathcal{G}(\mu).$$

Then $\mathcal{G}$ is totally convex.

Theorem (But not so many...)

Let $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ proper, lower semicontinuous and totally convex. Then

$$\mathcal{G}(\delta_{M(\mu)}) \leqslant \mathcal{G}(\mu) \quad \text{for every} \quad \mu \in \mathcal{P}(\mathbb{R}^d).$$

In particular, if $\mu \in \arg\min \mathcal{G}$ it also holds $\delta_{M(\mu)} \in \arg\min \mathcal{G}$.

Full result

Definition (Convex order on measures)

Given $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, μ and ν are in convex order $\mu \leqslant_{\mathsf{C}} \nu$ if $\int_{\mathbb{R}^d} f \mathrm{d}\mu \leqslant \int_{\mathbb{R}^d} f \mathrm{d}\nu$ for any f continuous, convex and in $L^1(\nu)$.

Theorem (Di Marino, Farinelli, N.)

Let $\mathcal{G} : \mathcal{P}_2(\mathbb{R}^d) \to (-\infty, +\infty]$ proper, lower semicontinuous and totally convex. Then for any $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ satisfying $\mu \leqslant_C \nu$, we have

$$\mathcal{G}(\mu) \leqslant \mathcal{G}(\nu).$$

Consequences [Di Marino, Farinelli, N.]

Lemma (Prox of deltas are deltas)

Let $\mathcal{G}$ be proper, lower semicontinuous and totally convex, then for every $x \in \mathbb{R}^d$ there exists y_x such that $\operatorname{prox}_{\tau\mathcal{G}}(\delta_x) = \delta_{y_x}$.

Consequences [Di Marino, Farinelli, N.]

Lemma (Prox of deltas are deltas)

Let $\mathcal{G}$ be proper, lower semicontinuous and totally convex, then for every $x \in \mathbb{R}^d$ there exists y_x such that $\mathrm{prox}_{\tau\mathcal{G}}(\delta_x) = \delta_{y_x}$.

Proposition

Let $\mathcal{G}$ be proper, lower semicontinuous and convex along (outer) generalized geodesics with $D(\mathcal{G}) \subset \mathcal{P}_2^r(\mathbb{R}^d)$. Then it cannot exist a sequence of totally convex functionals $\{\mathcal{G}_n\}_n$ such that $\mathrm{prox}_{\tau\mathcal{G}_n}(\mu) \to \mathrm{prox}_{\tau\mathcal{G}}(\mu)$ for every $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.

Consequences [Di Marino, Farinelli, N.]

Lemma (Prox of deltas are deltas)

Let $\mathcal{G}$ be proper, lower semicontinuous and totally convex, then for every $x \in \mathbb{R}^d$ there exists y_x such that $\mathrm{prox}_{\tau\mathcal{G}}(\delta_x) = \delta_{y_x}$.

Proposition

Let $\mathcal{G}$ be proper, lower semicontinuous and convex along (outer) generalized geodesics with $D(\mathcal{G}) \subset \mathcal{P}_2^r(\mathbb{R}^d)$. Then it cannot exist a sequence of totally convex functionals $\{\mathcal{G}_n\}_n$ such that $\mathrm{prox}_{\tau\mathcal{G}_n}(\mu) \to \mathrm{prox}_{\tau\mathcal{G}}(\mu)$ for every $\mu \in \mathcal{P}_2(\mathbb{R}^d)$.

Proposition

Let $\mathcal{G}^\tau$ be the Moreau envelope of $\mathcal{G}$ defined by

$$\mathcal{G}^\tau(\mu) := \inf_{\nu \in \mathcal{P}_2(\mathbb{R}^d)} \mathcal{G}(\nu) + \frac{1}{2\tau} W_2^2(\mu, \nu).$$

Let $\mathcal{G} : \mathcal{P}(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$ be the negative entropy. Then, the function $\mathcal{G}^\tau$ cannot be geodesically convex.

UniGe | MaLGa

Thank you for your attention!

References

Simone Di Marino, Emanuele Naldi, Silvia Villa.
Inexact JKO and proximal-gradient algorithms in the Wasserstein space.
arXiv preprint `arXiv:2505.23517`, 2025.

Simone Di Marino, Sara Farinelli, Emanuele Naldi.
Lipschitz properties of proximity operators in the Wasserstein space.
Work in progress (Just started!).