Gradient flows of potential energies in the geometry of Sinkhorn divergences

Hugo Lavenant

Bocconi University

Workshop “Wasserstein Gradient flows in Math and Machine Learning”

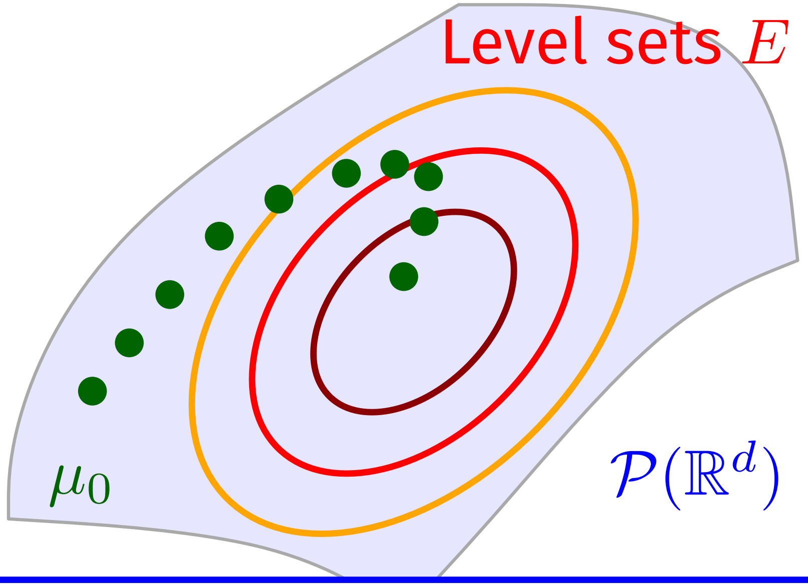
Banff (Canada), July 4, 2025

Joint work with

Mathis Hardion

Mathis' master thesis available online, preprint of the article on arxiv hopefully soon.

The JKO scheme



$E : \mathcal{P}(\mathbb{R}^d) \to [0, +\infty]$ and $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$.

$$\mathrm{OT}(\mu, \nu) = \min_{\pi \in \Pi(\mu,\nu)} \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \, \mathrm{d}\pi(x, y).$$

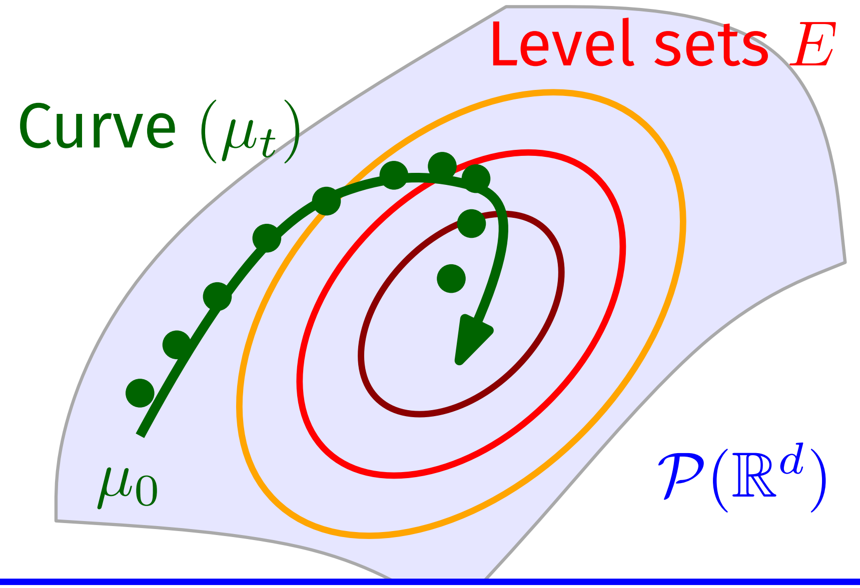
Subset of $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$, coupling between μ and ν

Minimizing movement scheme / JKO scheme. For $\tau > 0$ iterate for $k \geq 0$,

$$\mu_{k+1}^{\tau} \in \arg\min_{\mu} \; E(\mu) + \frac{\mathrm{OT}(\mu, \mu_k^{\tau})}{2\tau}$$

The JKO scheme

$E : \mathcal{P}(\mathbb{R}^d) \to [0, +\infty]$ and $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$.



$$\mathrm{OT}(\mu, \nu) = \min_{\pi \in \Pi(\mu,\nu)} \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \, \mathrm{d}\pi(x, y).$$

Subset of $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$, coupling between μ and ν

Minimizing movement scheme / JKO scheme. For $\tau > 0$ iterate for $k \geq 0$,

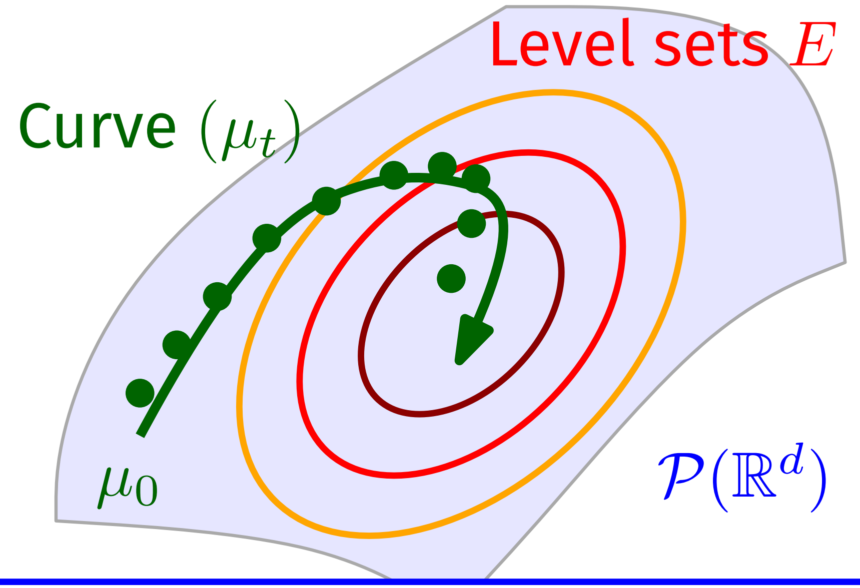
$$\mu_{k+1}^\tau \in \arg\min_{\mu} \; E(\mu) + \frac{\mathrm{OT}(\mu, \mu_k^\tau)}{2\tau}$$

$\mu_k^\tau \simeq \mu_{k\tau}$ as $\tau \to 0$ with $(\mu_t)_t$ **gradient flow** of E with respect to OT geometry.

[Jordan, Kinderlehrer & Otto, 1998]

The JKO scheme

$E : \mathcal{P}(\mathbb{R}^d) \to [0, +\infty]$ and $\mu_0 \in \mathcal{P}(\mathbb{R}^d)$.



$$\mathrm{OT}(\mu, \nu) = \min_{\pi \in \Pi(\mu,\nu)} \iint_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \, \mathrm{d}\pi(x, y).$$

Subset of $\mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$, coupling between μ and ν

Minimizing movement scheme / JKO scheme. For $\tau > 0$ iterate for $k \geq 0$,

$$\mu_{k+1}^\tau \in \arg\min_{\mu} \; E(\mu) + \frac{\mathrm{OT}(\mu, \mu_k^\tau)}{2\tau}$$

$\mu_k^\tau \simeq \mu_{k\tau}$ as $\tau \to 0$ with $(\mu_t)_t$ **gradient flow** of E with respect to OT geometry.

Question: how to use entropic OT instead of OT?

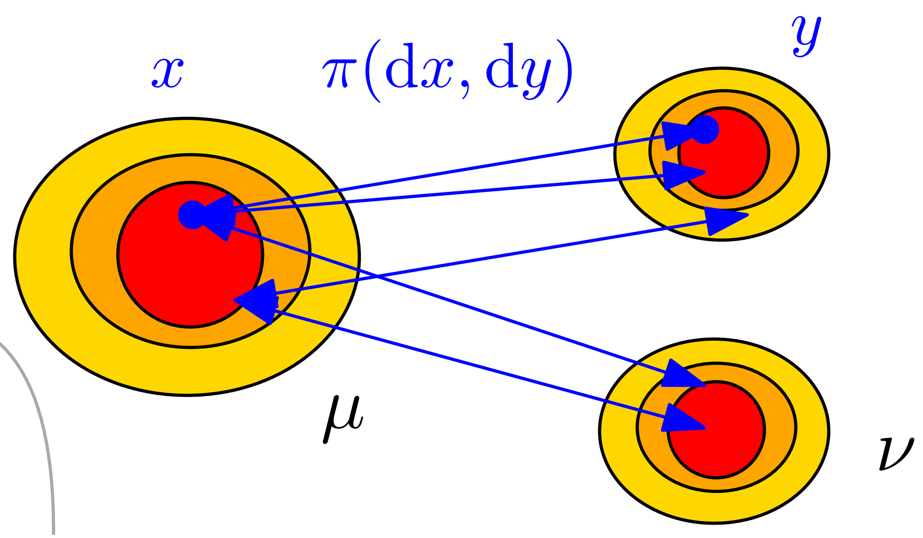
[Jordan, Kinderlehrer & Otto, 1998]

With entropic optimal transport?

(X, d) compact metric space with symmetric cost function c, and $\varepsilon > 0$.

Definition

$$\mathrm{OT}_\varepsilon(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \iint_{X\times X} c(x,y)\,\mathrm{d}\pi(x,y) + \varepsilon \mathrm{KL}(\pi|\mu\otimes\nu)$$



Kullback-Leibler divergence, a.k.a relative entropy

Why?

1. easier to compute (**Sinkhorn algorithm**),
2. better statistical complexity,
3. smoother dependence in (μ, ν).

[Peyré & Cuturi ,2018]

With entropic optimal transport?

(X, d) compact metric space with symmetric cost function c, and $\varepsilon > 0$.

Definition

$$\mathrm{OT}_\varepsilon(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \iint_{X\times X} c(x,y)\,\mathrm{d}\pi(x,y) + \varepsilon\mathrm{KL}(\pi|\mu\otimes\nu)$$

Entropic JKO scheme:

$$\mu^\tau_{k+1} \in \arg\min_\mu \; E(\mu) + \frac{\mathrm{OT}_\varepsilon(\mu,\mu^\tau_k)}{2\tau}$$

[Peyré, 2015]

With entropic optimal transport?

(X,d) compact metric space with symmetric cost function c, and $\varepsilon > 0$.

Definition

$$\mathrm{OT}_\varepsilon(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \iint_{X\times X} c(x,y)\,\mathrm{d}\pi(x,y) + \varepsilon \mathrm{KL}(\pi|\mu\otimes\nu)$$

Entropic JKO scheme:

$$\mu_{k+1}^\tau \in \arg\min_\mu \; E(\mu) + \frac{\mathrm{OT}_\varepsilon(\mu,\mu_k^\tau)}{2\tau}$$

Problem: μ does not minimize $\nu \mapsto \mathrm{OT}_\varepsilon(\nu,\mu)$.

$\rightsquigarrow$ Global minima of E not stable.

$\rightsquigarrow$ Temporal evolution not a time scale τ.

[Peyré, 2015]

With entropic optimal transport?

(X,d) compact metric space with symmetric cost function c, and $\varepsilon > 0$.

Definition

$$\mathrm{OT}_\varepsilon(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \iint_{X\times X} c(x,y)\,\mathrm{d}\pi(x,y) + \varepsilon \mathrm{KL}(\pi|\mu\otimes\nu)$$

Entropic JKO scheme:

$$\mu_{k+1}^\tau \in \arg\min_\mu \; E(\mu) + \frac{\mathrm{OT}_\varepsilon(\mu,\mu_k^\tau)}{2\tau}$$

Problem: μ does not minimize $\nu \mapsto \mathrm{OT}_\varepsilon(\nu,\mu)$.

$\rightsquigarrow$ Global minima of E not stable.

$\rightsquigarrow$ Temporal evolution not a time scale τ.

One Solution: Decrease ε with τ.

$\rightsquigarrow$ scaling $\varepsilon \ll \tau$ or $\varepsilon \sim \tau$.

[Carlier, Duval, Peyré, Schmitzer, 2017]
[Baradat, Hraivoronska, Santambrogio, 2025]

With entropic optimal transport?

(X,d) compact metric space with symmetric cost function c, and $\varepsilon > 0$.

Definition

$$\mathrm{OT}_\varepsilon(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \iint_{X\times X} c(x,y)\,\mathrm{d}\pi(x,y) + \varepsilon \mathrm{KL}(\pi|\mu\otimes\nu)$$

Entropic JKO scheme:

$$\mu_{k+1}^\tau \in \arg\min_\mu \; E(\mu) + \frac{\mathrm{OT}_\varepsilon(\mu,\mu_k^\tau)}{2\tau}$$

Problem: μ does not minimize $\nu \mapsto \mathrm{OT}_\varepsilon(\nu,\mu)$.

$\rightsquigarrow$ Global minima of E not stable.

$\rightsquigarrow$ Temporal evolution not a time scale τ.

One Solution: Decrease ε with τ.

$\rightsquigarrow$ scaling $\varepsilon \ll \tau$ or $\varepsilon \sim \tau$.

Today: I will keep ε fixed.

Using Sinkhorn divergences

As $\mathrm{OT}_\varepsilon(\mu,\mu) > 0$, **debias** by defining **Sinkhorn divergence**

$$S_\varepsilon(\mu,\nu) = \mathrm{OT}_\varepsilon(\mu,\nu) - \frac{1}{2}\mathrm{OT}_\varepsilon(\mu,\mu) - \frac{1}{2}\mathrm{OT}_\varepsilon(\nu,\nu).$$

[Genevay, Peyré, & Cuturi, 2018]

Using Sinkhorn divergences

As $\mathrm{OT}_\varepsilon(\mu,\mu) > 0$, **debias** by defining **Sinkhorn divergence**

$$S_\varepsilon(\mu,\nu) = \mathrm{OT}_\varepsilon(\mu,\nu) - \frac{1}{2}\mathrm{OT}_\varepsilon(\mu,\mu) - \frac{1}{2}\mathrm{OT}_\varepsilon(\nu,\nu).$$

Assumption until the end of the talk

Theorem Assume $\exp(-c/\varepsilon)$ positive definite universal kernel.

1. $S_\varepsilon(\mu,\nu) \geq 0$ with equality iff $\mu = \nu$, and S_ε "metrizes" weak convergence.
2. S_ε convex in each of its inputs.

but $\sqrt{S_\varepsilon}$ not a distance

[Genevay, Peyré, & Cuturi, 2018]
[Feydy, Séjourné, Vialard, Amari, Trouvé & Peyré, 2019]

Using Sinkhorn divergences

As $\mathrm{OT}_\varepsilon(\mu,\mu) > 0$, **debias** by defining **Sinkhorn divergence**

$$S_\varepsilon(\mu,\nu) = \mathrm{OT}_\varepsilon(\mu,\nu) - \frac{1}{2}\mathrm{OT}_\varepsilon(\mu,\mu) - \frac{1}{2}\mathrm{OT}_\varepsilon(\nu,\nu).$$

Assumption until the end of the talk

Theorem Assume $\exp(-c/\varepsilon)$ positive definite universal kernel.

1. $S_\varepsilon(\mu,\nu) \geq 0$ with equality iff $\mu = \nu$, and S_ε "metrizes" weak convergence.
2. S_ε convex in each of its inputs.

but $\sqrt{S_\varepsilon}$ not a distance

Sinkhorn JKO:

$$\mu_{k+1}^\tau \in \arg\min_\mu \; E(\mu) + \frac{S_\varepsilon(\mu,\mu_k^\tau)}{2\tau}$$

See also [Aubin-Frankowski, Sodini & Stefanelli, 2025]

Using Sinkhorn divergences

As $\mathrm{OT}_\varepsilon(\mu,\mu) > 0$, **debias** by defining **Sinkhorn divergence**

$$S_\varepsilon(\mu,\nu) = \mathrm{OT}_\varepsilon(\mu,\nu) - \frac{1}{2}\mathrm{OT}_\varepsilon(\mu,\mu) - \frac{1}{2}\mathrm{OT}_\varepsilon(\nu,\nu).$$

Assumption until the end of the talk

Theorem Assume $\exp(-c/\varepsilon)$ positive definite universal kernel.

1. $S_\varepsilon(\mu,\nu) \geq 0$ with equality iff $\mu=\nu$, and S_ε "metrizes" weak convergence.
2. S_ε convex in each of its inputs.

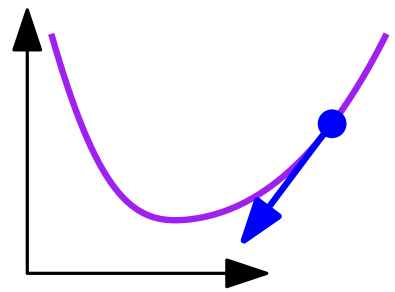
but $\sqrt{S_\varepsilon}$ not a distance

Sinkhorn JKO:

$$\mu_{k+1}^\tau \in \arg\min_\mu \; E(\mu) + \frac{S_\varepsilon(\mu,\mu_k^\tau)}{2\tau}$$

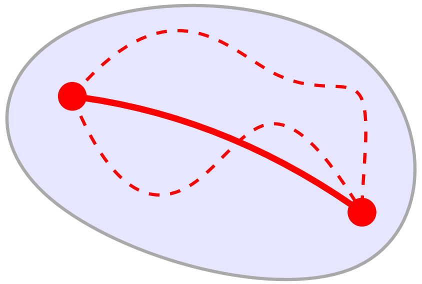
Today: Study of SJKO with $E(\mu) = \int V \,\mathrm{d}\mu$ potential energy.

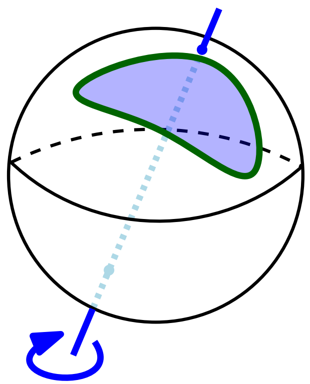
(if $\varepsilon = 0$ convergence to the transport equation $\partial_t \mu = \mathrm{div}(\mu \nabla V)$)



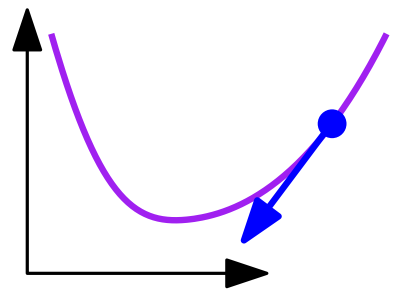
1 - Derivation of the limit flow as $\tau \to 0$

2 - Interlude: the Riemannian geometry of Sinkhorn divergences



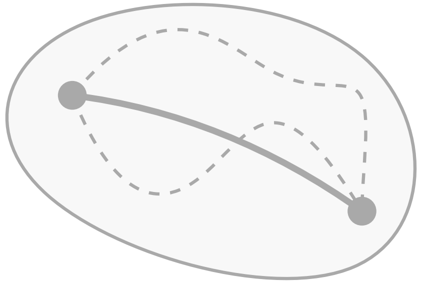


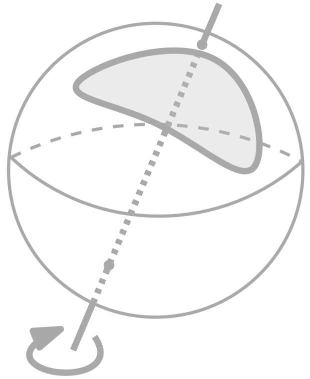
3 - Existence, stability and long term behavior of the flow



1 - Derivation of the limit flow as $\tau \to 0$

2 - Interlude: the Riemannian geometry of Sinkhorn divergences





3 - Existence, stability and long term behavior of the flow

Technical preliminaries: Schrödinger potentials

$(f_{\mu,\nu}, g_{\mu,\nu})$ **Schrödinger potentials**, solutions of: $\begin{cases} f = T_\varepsilon(g,\nu), \\ g = T_\varepsilon(f,\mu), \end{cases}$

with $T_\varepsilon(f,\mu) = -\varepsilon \log \int_X \exp\left(\frac{f(x) - c(x,\cdot)}{\varepsilon}\right) \mathrm{d}\mu(x).$

Proposition

$$\frac{\delta \mathrm{OT}_\varepsilon(\mu,\nu)}{\delta(\mu,\nu)} = (f_{\mu,\nu}, g_{\mu,\nu}).$$

[Feydy, Séjourné, Vialard, Amari, Trouvé & Peyré, 2019]

Technical preliminaries: Schrödinger potentials

$(f_{\mu,\nu}, g_{\mu,\nu})$ **Schrödinger potentials**, solutions of: $\begin{cases} f = T_\varepsilon(g,\nu), \\ g = T_\varepsilon(f,\mu), \end{cases}$

with $T_\varepsilon(f,\mu) = -\varepsilon \log \int_X \exp\left(\frac{f(x) - c(x,\cdot)}{\varepsilon}\right) \mathrm{d}\mu(x).$

Proposition

$$\frac{\delta \mathrm{OT}_\varepsilon(\mu,\nu)}{\delta(\mu,\nu)} = (f_{\mu,\nu}, g_{\mu,\nu}).$$

Definition:

$$k_{\mu,\nu}(x,y) = \exp\left(\frac{f_{\mu,\nu}(x) + g_{\mu,\nu}(y) - c(x,y)}{\varepsilon}\right).$$

π_ε entropic optimal plan between μ and ν is:

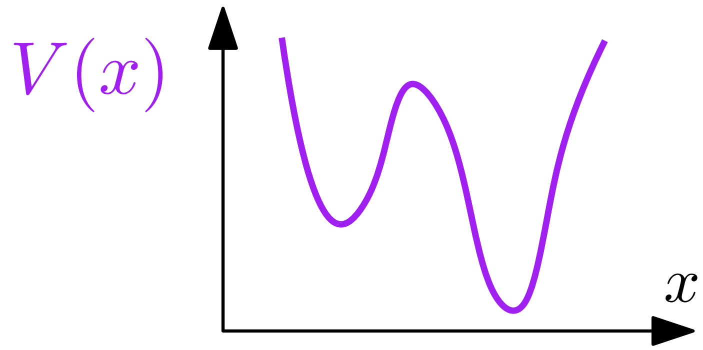
$$\pi_\varepsilon = k_{\mu,\nu} \cdot (\mu \otimes \nu).$$

Optimality conditions of SJKO

Sinkhorn JKO:

$$\mu_{k+1}^{\tau} \in \arg\min_{\mu} \; E(\mu) + \frac{S_\varepsilon(\mu, \mu_k^\tau)}{2\tau}.$$

with S_ε Sinkhorn divergence and $E(\mu) = \int V \,\mathrm{d}\mu$ **potential energy**.



Optimality conditions of SJKO

Sinkhorn JKO:

$$\mu_{k+1}^\tau \in \arg\min_\mu \; E(\mu) + \frac{S_\varepsilon(\mu, \mu_k^\tau)}{2\tau}.$$

with S_ε Sinkhorn divergence and $E(\mu) = \int V \,\mathrm{d}\mu$ **potential energy**.

Lemma. For any k, there exists $p_k^\tau \leq 0$ with $p_k^\tau = 0$ on $\mathrm{supp}(\mu_{k+1}^\tau)$ s.t.

$$\frac{f_{\mu_{k+1}^\tau, \mu_k^\tau} - f_{\mu_{k+1}^\tau, \mu_{k+1}^\tau}}{2\tau} + V + p_k^\tau = \text{Cst}.$$

Optimality conditions of SJKO

Sinkhorn JKO:

$$\mu_{k+1}^\tau \in \arg\min_\mu \; E(\mu) + \frac{S_\varepsilon(\mu, \mu_k^\tau)}{2\tau}.$$

with S_ε Sinkhorn divergence and $E(\mu) = \int V \, \mathrm{d}\mu$ **potential energy**.

Lemma. For any k, there exists $p_k^\tau \leq 0$ with $p_k^\tau = 0$ on $\mathrm{supp}(\mu_{k+1}^\tau)$ s.t.

$$\frac{f_{\mu_{k+1}^\tau, \mu_k^\tau} - f_{\mu_{k+1}^\tau, \mu_{k+1}^\tau}}{2\tau} + V + p_k^\tau = \text{Cst}.$$

Must be related to $\mu_{k+1}^\tau - \mu_k^\tau$

Equation of the limit flow

Define

$$K_\mu[\phi](x) = \int_X k_{\mu,\mu}(x,y)\phi(y)\,\mathrm{d}\mu(y),$$

$$H_\mu[\sigma](x) = \int_X k_{\mu,\mu}(x,y)\,\mathrm{d}\sigma(y).$$

Theorem (informal)

$$\left.\frac{\partial f_{\mu,\nu}}{\partial \nu}\right|_{\mu=\nu}[\sigma] = -\varepsilon(\mathrm{Id} - K_\mu^2)^{-1} H_\mu[\sigma].$$

[Carlier & Laborde, 2020], [Sanz, Loubes, and Niles-Weed, 2022], [Lavenant, Luckhardt, Mordant, Schmitzer, Tamanini, 2024]

Equation of the limit flow

Define

$$K_\mu[\phi](x) = \int_X k_{\mu,\mu}(x,y)\phi(y)\,\mathrm{d}\mu(y),$$

$$H_\mu[\sigma](x) = \int_X k_{\mu,\mu}(x,y)\,\mathrm{d}\sigma(y).$$

Theorem (informal)

$$\left.\frac{\partial f_{\mu,\nu}}{\partial \nu}\right|_{\mu=\nu}[\sigma] = -\varepsilon(\mathrm{Id} - K_\mu^2)^{-1} H_\mu[\sigma].$$

Recall: SJKO yields $\dfrac{f_{\mu^\tau_{k+1},\mu^\tau_k} - f_{\mu^\tau_{k+1},\mu^\tau_{k+1}}}{2\tau} + V + p^\tau_k =$ Cst.

[Carlier & Laborde, 2020], [Sanz, Loubes, and Niles-Weed, 2022], [Lavenant, Luckhardt, Mordant, Schmitzer, Tamanini, 2024]

Equation of the limit flow

Define

$$K_\mu[\phi](x) = \int_X k_{\mu,\mu}(x,y)\phi(y)\,\mathrm{d}\mu(y),$$

$$H_\mu[\sigma](x) = \int_X k_{\mu,\mu}(x,y)\,\mathrm{d}\sigma(y).$$

Theorem (informal)

$$\left.\frac{\partial f_{\mu,\nu}}{\partial \nu}\right|_{\mu=\nu}[\sigma] = -\varepsilon(\mathrm{Id} - K_\mu^2)^{-1}H_\mu[\sigma].$$

Formal limit $\tau \to 0$ of SJKO: **Sinkhorn potential flow**

$$\frac{\varepsilon}{2}(\mathrm{Id} - K_{\mu_t}^2)^{-1}H_{\mu_t}[\dot{\mu}_t] + V + p_t = \text{Cst}.$$

p_t pressure: $p_t \leq 0$, and $p_t = 0$ on $\mathrm{supp}(\mu_t)$

Equation of the limit flow

Define

$$K_\mu[\phi](x) = \int_X k_{\mu,\mu}(x,y)\phi(y)\,\mathrm{d}\mu(y),$$

$$H_\mu[\sigma](x) = \int_X k_{\mu,\mu}(x,y)\,\mathrm{d}\sigma(y).$$

Theorem (informal)

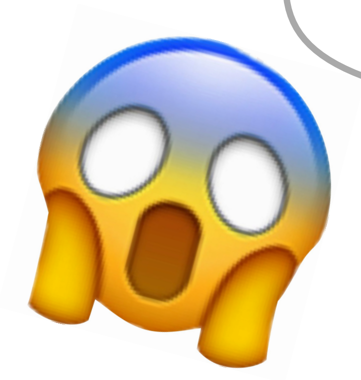
$$\left.\frac{\partial f_{\mu,\nu}}{\partial \nu}\right|_{\mu=\nu}[\sigma] = -\varepsilon(\mathrm{Id} - K_\mu^2)^{-1} H_\mu[\sigma].$$

Formal limit $\tau \to 0$ of SJKO: **Sinkhorn potential flow**

$$\frac{\varepsilon}{2}(\mathrm{Id} - K_{\mu_t}^2)^{-1} H_{\mu_t}[\dot{\mu}_t] + V + p_t = \text{Cst.}$$

p_t pressure: $p_t \leq 0$, and $p_t = 0$ on $\mathrm{supp}(\mu_t)$

So $\dot{\mu}_t = H_{\mu_t}^{-1}[\ldots]$ with H_{μ_t} "convolution".
Non local equation **of infinite order**.

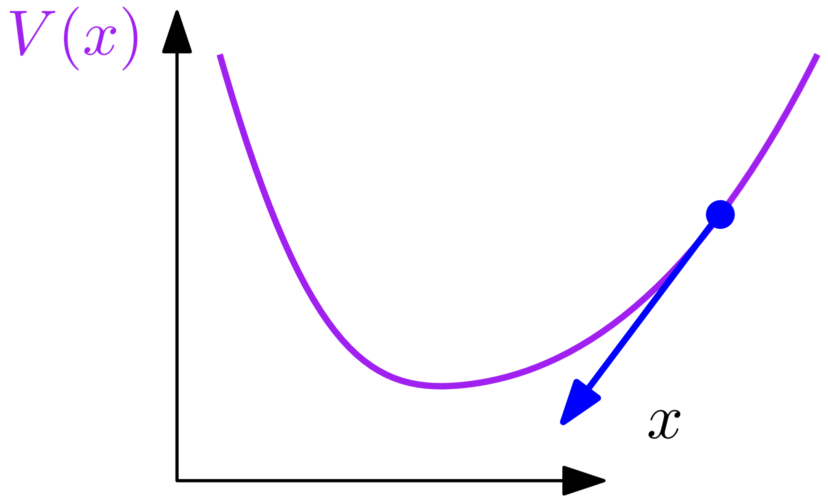


Examples

Proposition. If $X = \mathbb{R}^d$, V convex, c quadratic cost and $\mu_0 = \delta_{x_0}$ then $\mu_t = \delta_{x_t}$ with

$$\dot{x}_t \in -\partial V(x_t).$$

(Same as Wasserstein GF).

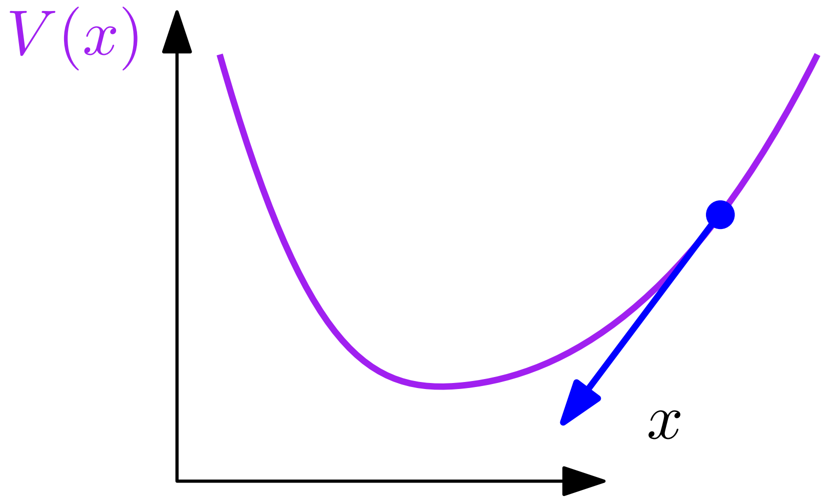


Examples

Proposition. If $X = \mathbb{R}^d$, V convex, c quadratic cost and $\mu_0 = \delta_{x_0}$ then $\mu_t = \delta_{x_t}$ with

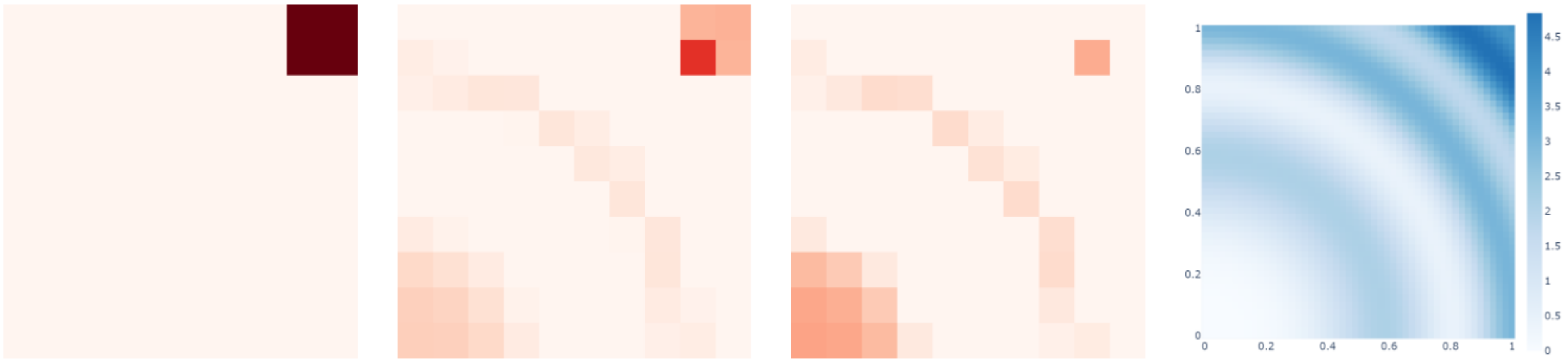
$$\dot{x}_t \in -\partial V(x_t).$$

(Same as Wasserstein GF).

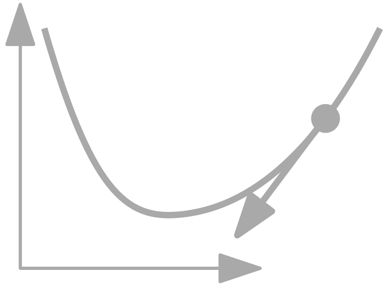


But if V not convex there can be **teleportation**!

Here a **Lagrangian** discretization won't work.

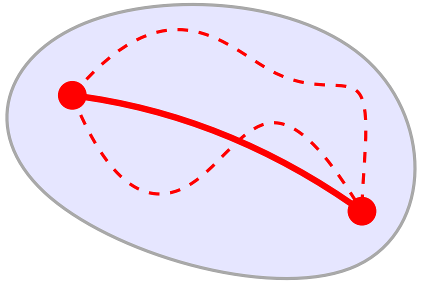


$t = 0.00$ $t = 0.25$ $t = 0.50$ V



1 - Derivation of the limit flow as $\tau \to 0$

2 - Interlude: the Riemannian geometry of Sinkhorn divergences



Joint work with:

J. Luckhardt

G. Mordant

B. Schmitzer

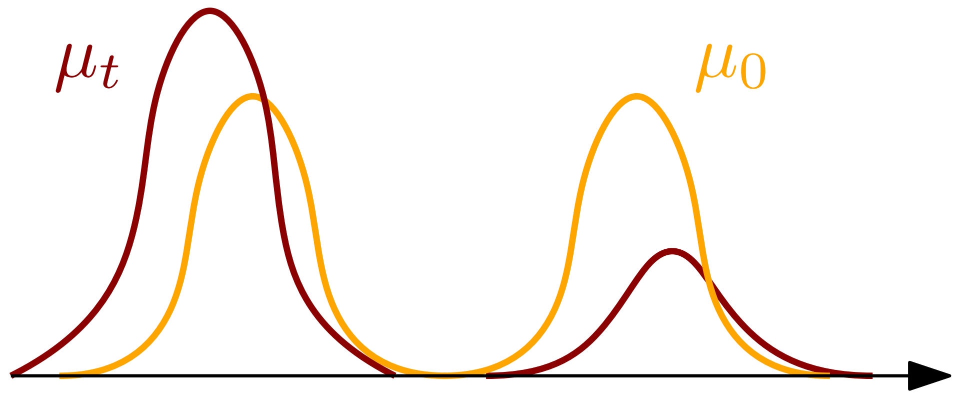
L. Tamanini

The Hessian of the Sinkhorn divergence

$(\mu_t)_t$ curve with $\dot{\mu}_t \in C^m(X)^*$ and $c \in C^{2m}(X \times X)$.

Theorem.

$$S_\varepsilon(\mu_0, \mu_t) \sim t^2 \frac{\varepsilon}{2} \langle \dot{\mu}, (\mathrm{Id} - K_\mu^2)^{-1} H_\mu[\dot{\mu}] \rangle.$$



Recall

$$K_\mu[\phi](x) = \int_X k_{\mu,\mu}(x,y)\phi(y)\,\mathrm{d}\mu(y),$$

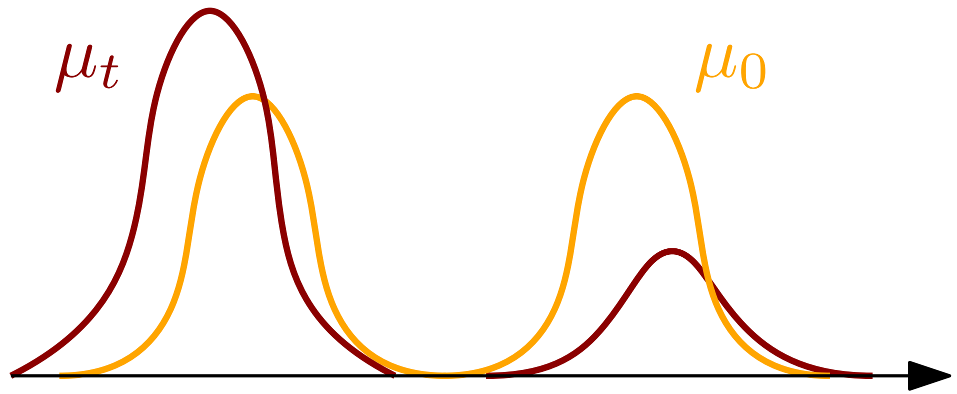
$$H_\mu[\sigma](x) = \int_X k_{\mu,\mu}(x,y)\,\mathrm{d}\sigma(y).$$

The Hessian of the Sinkhorn divergence

$(\mu_t)_t$ curve with $\dot\mu_t \in C^m(X)^*$ and $c \in C^{2m}(X \times X)$.

Theorem.

$$S_\varepsilon(\mu_0, \mu_t) \sim t^2 \frac{\varepsilon}{2} \langle \dot\mu, (\mathrm{Id} - K_\mu^2)^{-1} H_\mu[\dot\mu] \rangle.$$

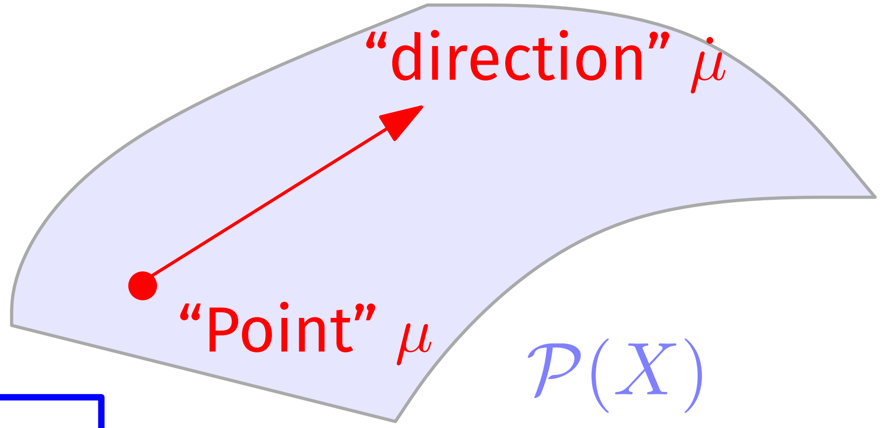


Same formula

Recall

$$K_\mu[\phi](x) = \int_X k_{\mu,\mu}(x,y)\phi(y)\,\mathrm{d}\mu(y),$$

$$H_\mu[\sigma](x) = \int_X k_{\mu,\mu}(x,y)\,\mathrm{d}\sigma(y).$$



Definition. $\mathrm{g}_\mu(\dot\mu, \dot\mu) = \frac{\varepsilon}{2} \langle \dot\mu, (\mathrm{Id} - K_\mu^2)^{-1} H_\mu[\dot\mu] \rangle.$

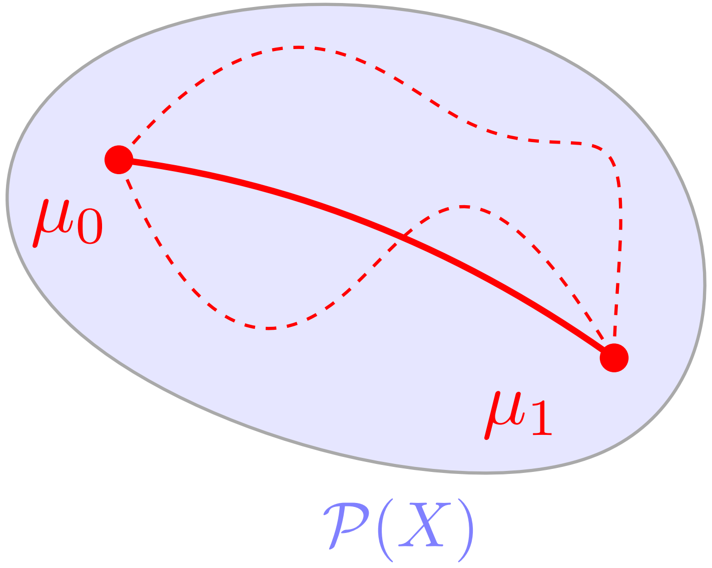
Definition of the Riemannian distance

Recall X compact, $\mathbf{g}_\mu(\dot\mu,\dot\mu) = \frac{\varepsilon}{2}\langle \dot\mu, (\mathrm{Id} - K_\mu^2)^{-1} H_\mu[\dot\mu]\rangle$.

Definition. Given μ_0, μ_1:

$$\mathrm{d}_S(\mu_0,\mu_1)^2 = \inf \int_0^1 \mathbf{g}_{\mu_t}(\dot\mu_t,\dot\mu_t)\,\mathrm{d}t$$

where infimum over $(\mu_t)_t$ in a suitable class of paths.



Both “vertical” and “horizontal” motion is allowed!

[Lavenant, Luckhardt, Mordant, Schmitzer, Tamanini, 2024]

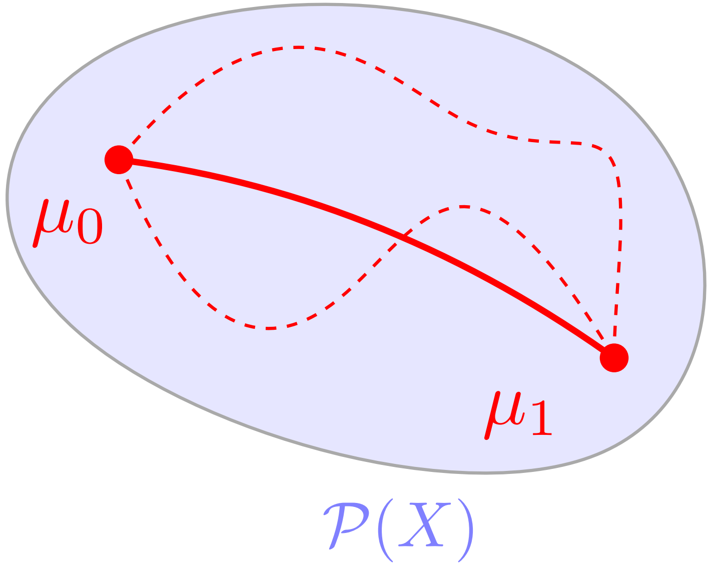
Definition of the Riemannian distance

Recall X compact, $\mathbf{g}_\mu(\dot\mu, \dot\mu) = \frac{\varepsilon}{2}\langle \dot\mu, (\mathrm{Id} - K_\mu^2)^{-1} H_\mu[\dot\mu] \rangle$.

Definition. Given μ_0, μ_1:

$$\mathrm{d}_S(\mu_0, \mu_1)^2 = \inf \int_0^1 \mathbf{g}_{\mu_t}(\dot\mu_t, \dot\mu_t)\,\mathrm{d}t$$

where infimum over $(\mu_t)_t$ in a suitable class of paths.



Both "vertical" and "horizontal" motion is allowed!

Theorem. d_S is a distance over $\mathcal{P}(X)$ **metrizing weak convergence of measures**, and the infimum in the definition is reached (**geodesics exist**).

[Lavenant, Luckhardt, Mordant, Schmitzer, Tamanini, 2024]

Link with the Sinkhorn potential flow

Sinkhorn potential flow

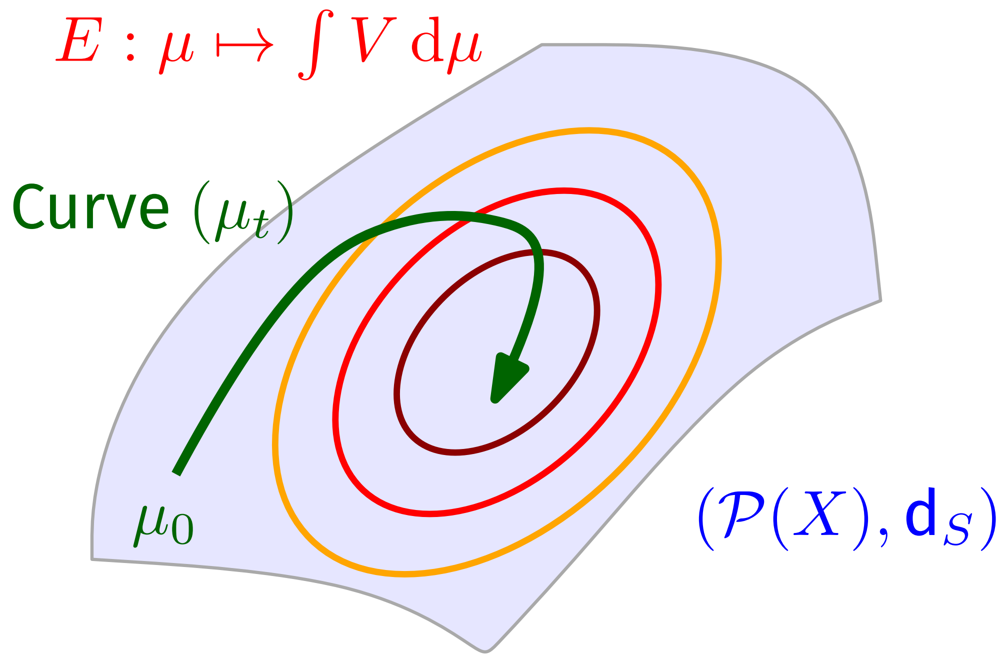
$$\frac{\varepsilon}{2}(\mathrm{Id} - K_{\mu_t}^2)^{-1} H_{\mu_t}[\dot{\mu}_t] + V + p_t = \text{Cst}.$$

Metric tensor

$$\mathbf{g}_\mu(\dot{\mu}, \dot{\mu}) = \frac{\varepsilon}{2}\langle \dot{\mu}, (\mathrm{Id} - K_\mu^2)^{-1} H_\mu[\dot{\mu}]\rangle.$$

Distance

$$\mathsf{d}_S(\mu_0, \mu_1)^2 = \inf \int_0^1 \mathbf{g}_{\mu_t}(\dot{\mu}_t, \dot{\mu}_t)\,\mathrm{d}t.$$



Link with the Sinkhorn potential flow

Sinkhorn potential flow

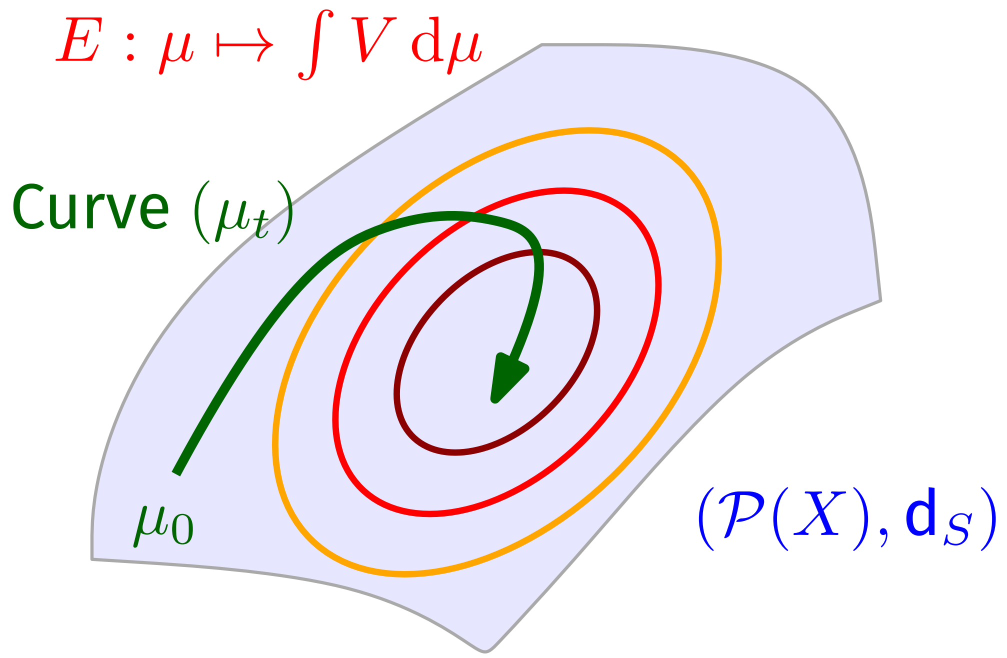
$$\frac{\varepsilon}{2}(\mathrm{Id} - K_{\mu_t}^2)^{-1} H_{\mu_t}[\dot\mu_t] + V + p_t = \text{Cst.}$$

Metric tensor

$$\mathbf{g}_\mu(\dot\mu, \dot\mu) = \frac{\varepsilon}{2}\langle \dot\mu, (\mathrm{Id} - K_\mu^2)^{-1} H_\mu[\dot\mu]\rangle.$$

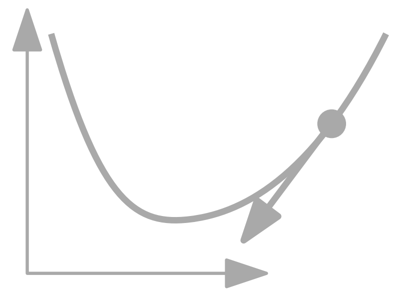
Distance

$$\mathsf{d}_S(\mu_0, \mu_1)^2 = \inf \int_0^1 \mathbf{g}_{\mu_t}(\dot\mu_t, \dot\mu_t)\,\mathrm{d}t.$$



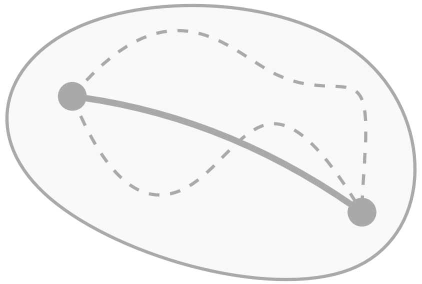
Claim: the Sinkhorn potential flow is formally the **gradient flow** of E in the Riemannian manifold $(\mathcal{P}(X), \mathbf{g})$.

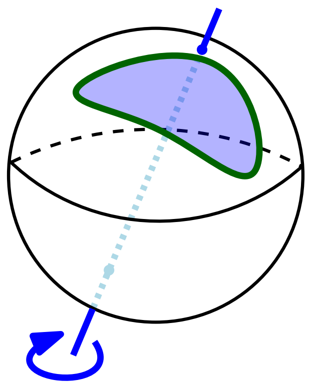
e.g. $\mathbf{g}_{\mu_t}(\dot\mu_t, z) + DE(\mu_t)(z) \geq 0$ for all $z \in T_{\mu_t}\mathcal{P}(X)$ admissible tangent vector.



1 - Derivation of the limit flow as $\tau \to 0$

2 - Interlude: the Riemannian geometry of Sinkhorn divergences





3 - Existence, stability and long term behavior of the flow

Main theoretical results

Sinkhorn potential flow: $\frac{\varepsilon}{2}(\mathrm{Id} - K_{\mu_t}^2)^{-1} H_{\mu_t}[\dot{\mu}_t] + V + p_t =$ Cst.

Theorem. (X compact, $\exp(-c/\varepsilon)$ p.d. universal, V continuous)

Main theoretical results

Sinkhorn potential flow: $\frac{\varepsilon}{2}(\mathrm{Id} - K_{\mu_t}^2)^{-1} H_{\mu_t}[\dot{\mu}_t] + V + p_t =$ Cst.

Theorem. (X compact, $\exp(-c/\varepsilon)$ p.d. universal, V continuous)

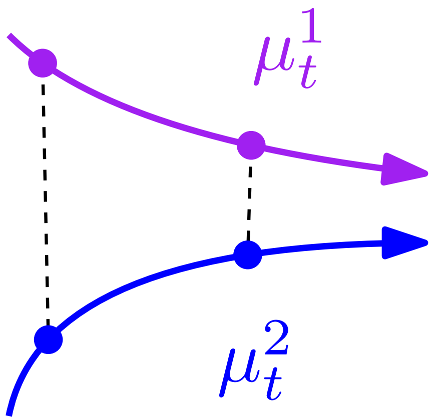
1. **Existence**: for any μ_0, there exists a global solution $(\mu_t)_{t\geq 0}$.

Main theoretical results

Sinkhorn potential flow: $\frac{\varepsilon}{2}(\mathrm{Id} - K_{\mu_t}^2)^{-1} H_{\mu_t}[\dot{\mu}_t] + V + p_t = \text{Cst.}$

Theorem. (X compact, $\exp(-c/\varepsilon)$ p.d. universal, V continuous)

1. **Existence**: for any μ_0, there exists a global solution $(\mu_t)_{t\geq 0}$.
2. The flow is **non-expansive** for a distance, equivalent to d_S, to be specified later.



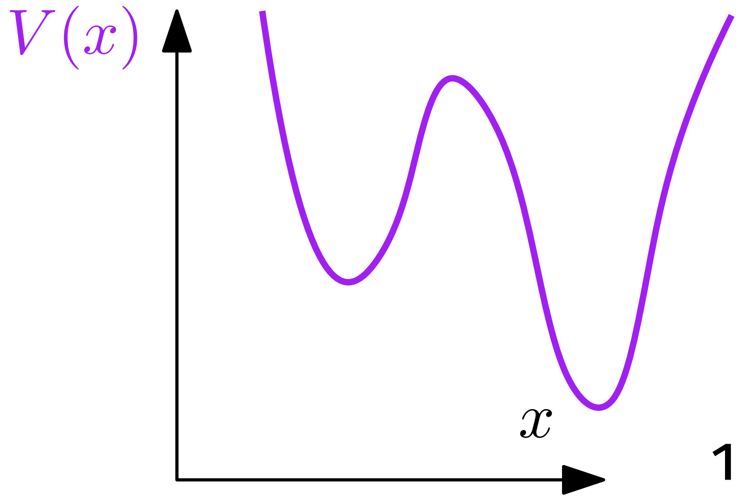
Main theoretical results

Sinkhorn potential flow: $\frac{\varepsilon}{2}(\mathrm{Id} - K^2_{\mu_t})^{-1} H_{\mu_t}[\dot{\mu}_t] + V + p_t = $ Cst.

Theorem. (X compact, $\exp(-c/\varepsilon)$ p.d. universal, V continuous)

1. **Existence**: for any μ_0, there exists a global solution $(\mu_t)_{t\geq 0}$.
2. The flow is **non-expansive** for a distance, equivalent to d_S, to be specified later.
3. Convergence to **global minimum**: $E(\mu_t) \to \min E$ as $t \to +\infty$.

Recall. The flow of the Wasserstein GF $\partial \mu_t = \mathrm{div}(\mu_t \nabla V)$ gets trapped in local minima.



The key change of variables into a Reproducing Kernel Hilbert Space

Recall $k_c = \exp(-c/\varepsilon) : X \times X \to \mathbb{R}$ positive definite and universal.

Definition. $\mathcal{H}_c$ Hilbert space of functions $X \to \mathbb{R}$: completion of

$$\operatorname{span}\{k_c(\cdot, x) \ : \ x \in X\}$$

with $\langle k_c(\cdot, x), k_c(\cdot, y)\rangle_{\mathcal{H}_c} = k_c(x, y)$.

k_c positive definite if this defines dot product

(k_c universal $\Leftrightarrow \mathcal{H}_k$ dense in $C(X)$)

[Paulsen & Raghupathi, 2016]

The key change of variables into a Reproducing Kernel Hilbert Space

Recall $k_c = \exp(-c/\varepsilon) : X \times X \to \mathbb{R}$ positive definite and universal.

Definition. $\mathcal{H}_c$ Hilbert space of functions $X \to \mathbb{R}$: completion of

$$\mathrm{span}\,\{k_c(\cdot, x) \ : \ x \in X\}$$

with $\langle k_c(\cdot, x), k_c(\cdot, y)\rangle_{\mathcal{H}_c} = k_c(x, y)$.

Define:

$$b = B(\mu) = \exp\left(-\frac{f_{\mu,\mu}}{\varepsilon}\right)$$

where $f_{\mu,\mu} : X \to \mathbb{R}$ Schrödinger potential.

[Feydy, Séjourné, Vialard, Amari, Trouvé & Peyré, 2019]

The key change of variables into a Reproducing Kernel Hilbert Space

Recall $k_c = \exp(-c/\varepsilon) : X \times X \to \mathbb{R}$ positive definite and universal.

Definition. $\mathcal{H}_c$ Hilbert space of functions $X \to \mathbb{R}$: completion of

$$\mathrm{span}\,\{k_c(\cdot, x) \ : \ x \in X\}$$

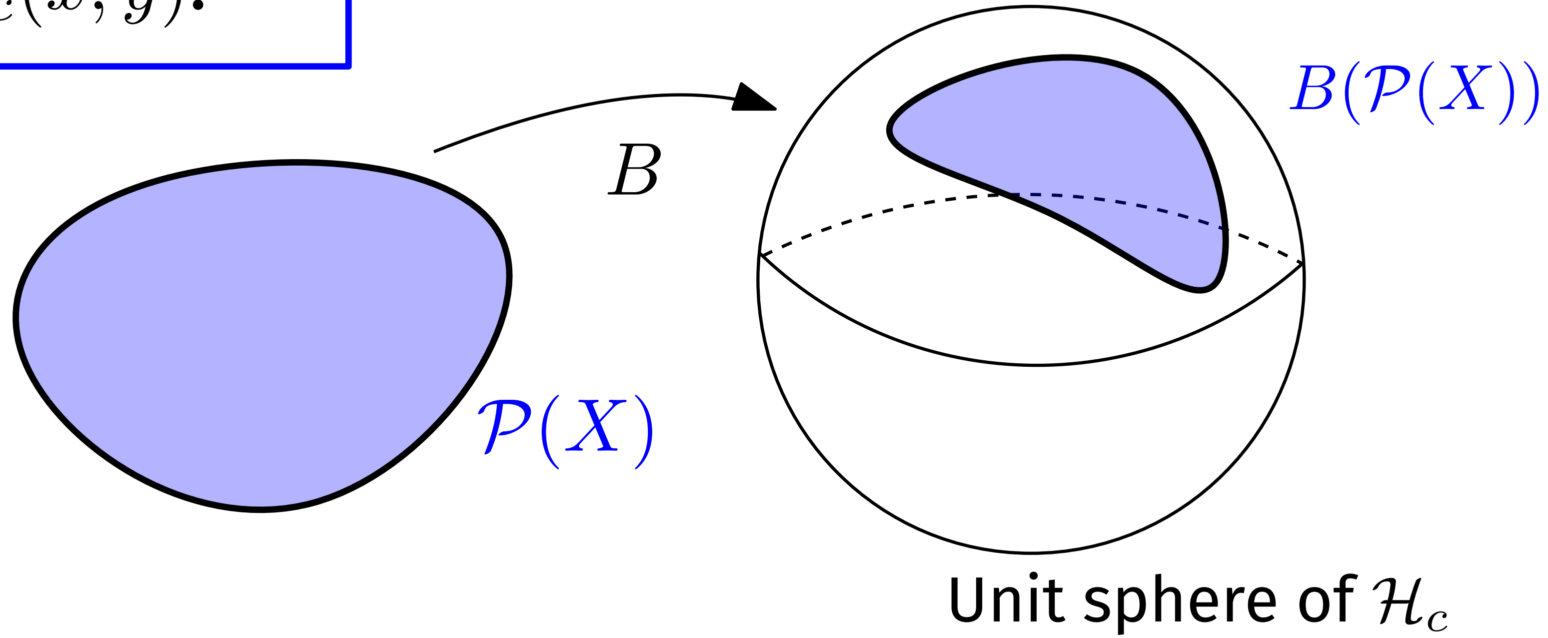
with $\langle k_c(\cdot, x), k_c(\cdot, y)\rangle_{\mathcal{H}_c} = k_c(x, y)$.

Define:

$$b = B(\mu) = \exp\left(-\frac{f_{\mu,\mu}}{\varepsilon}\right)$$

where $f_{\mu,\mu} : X \to \mathbb{R}$ Schrödinger potential.

Theorem. The map B is an homeomorphism between $\mathcal{P}(X)$ and the intersection of a convex cone and the unit sphere of $\mathcal{H}_c$.



[Lavenant, Luckhardt, Mordant, Schmitzer, & Tamanini, 2024]

Sinkhorn flow in the Hilbert space $\mathcal{H}_c$

With $b_t = \exp(-f_{\mu_t,\mu_t}/\varepsilon) \in \mathcal{H}_c$ and $V : X \to \mathbb{R}$.

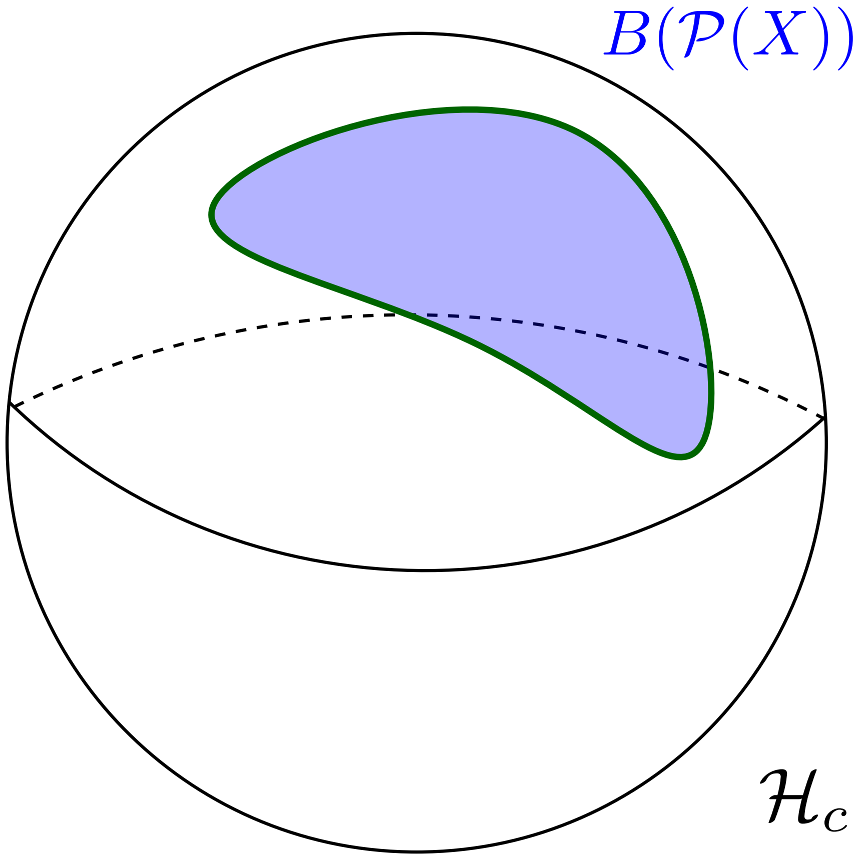
Sinkhorn flow in the b-variable.

$$\dot{b}_t + \frac{2}{\varepsilon}(V - V^*)\, b_t + p_t = 0.$$

$p_t \leq 0$, and $p_t = 0$ on $\mathrm{supp}(\mu_t)$

multiplication by V in $\mathcal{H}_c$

V^* Adjoint of multiplication by V for $\langle \cdot, \cdot \rangle_{\mathcal{H}_c}$



Sinkhorn flow in the Hilbert space $\mathcal{H}_c$

With $b_t = \exp(-f_{\mu_t,\mu_t}/\varepsilon) \in \mathcal{H}_c$ and $V : X \to \mathbb{R}$.

Sinkhorn flow in the b-variable.

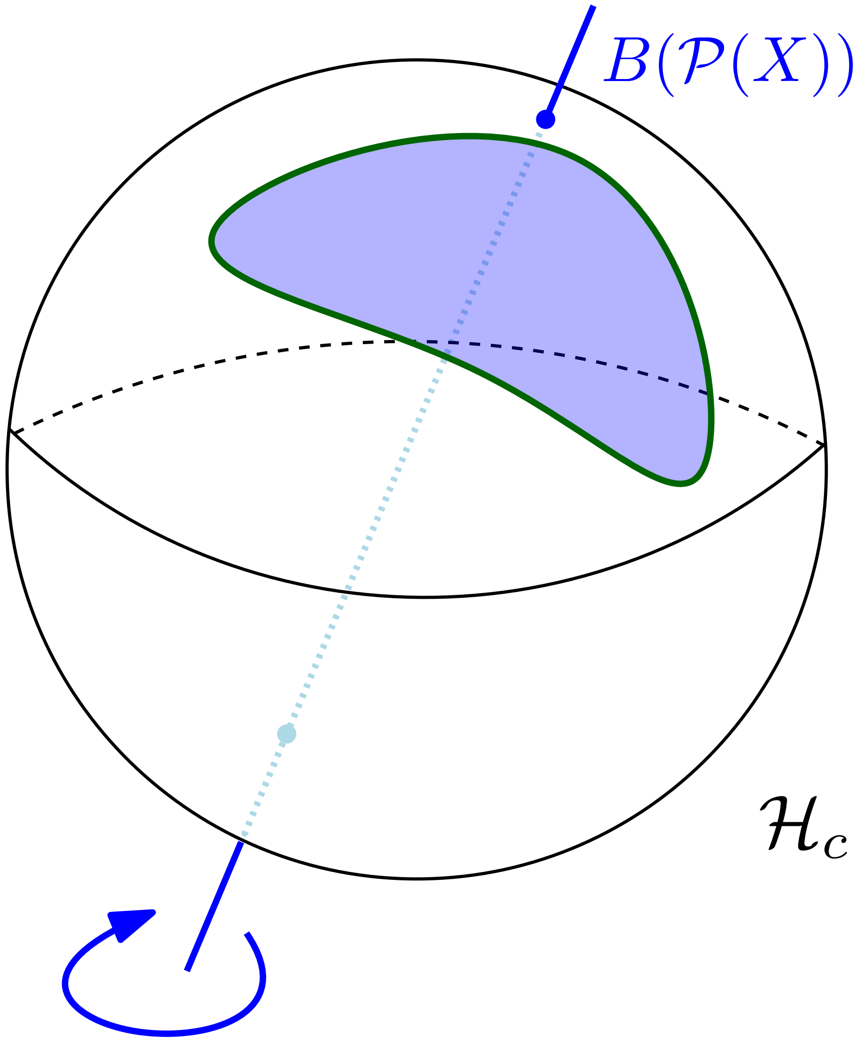
$$\dot{b}_t + \frac{2}{\varepsilon}(V - V^*)\, b_t + p_t = 0.$$

$p_t \leq 0$, and $p_t = 0$ on $\mathrm{supp}(\mu_t)$

multiplication by V in $\mathcal{H}_c$

V^* Adjoint of multiplication by V for $\langle \cdot, \cdot \rangle_{\mathcal{H}_c}$

$\frac{2}{\varepsilon}(V - V^*)$ skew-symmetric: generates group of unitary operators, but unbounded.



“Rotation” generated by $\frac{2}{\varepsilon}(V - V^*)$

Sinkhorn flow in the Hilbert space $\mathcal{H}_c$

With $b_t = \exp(-f_{\mu_t,\mu_t}/\varepsilon) \in \mathcal{H}_c$ and $V : X \to \mathbb{R}$.

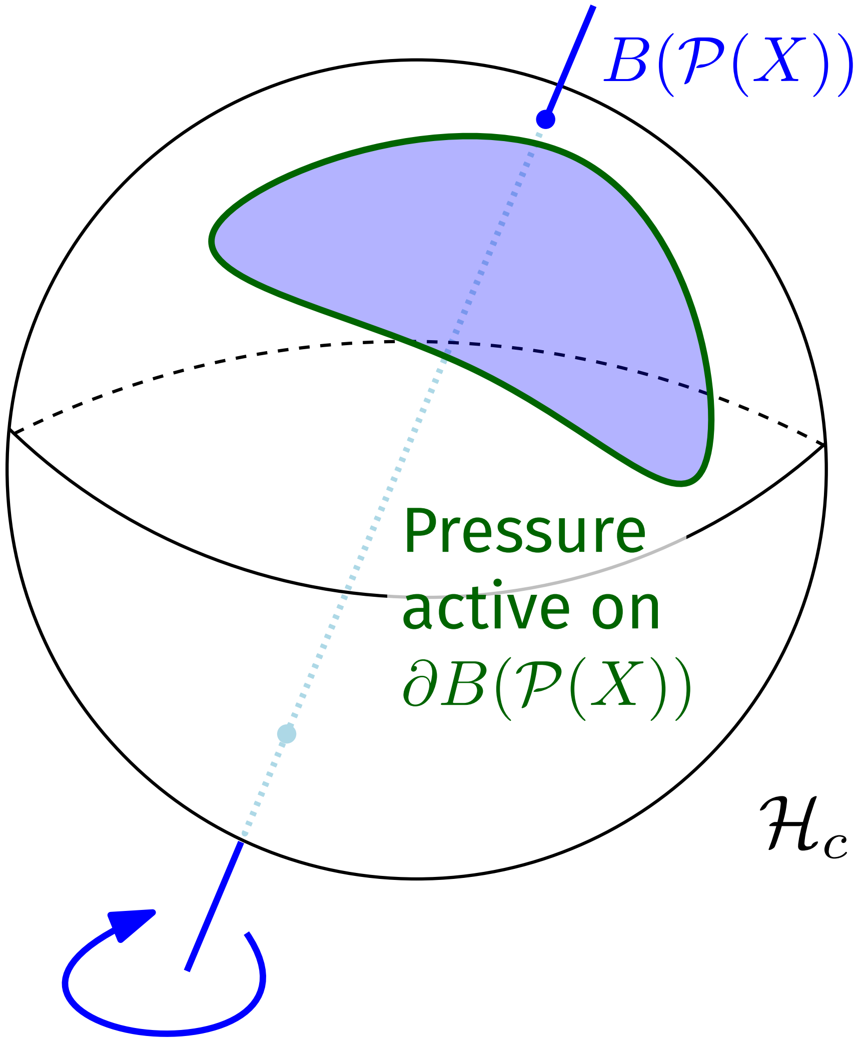
Sinkhorn flow in the b-variable.

$$\dot{b}_t + \frac{2}{\varepsilon}(V - V^*)\, b_t + p_t = 0.$$

$p_t \leq 0$, and $p_t = 0$ on $\mathrm{supp}(\mu_t)$

multiplication by V in $\mathcal{H}_c$

V^* Adjoint of multiplication by V for $\langle \cdot, \cdot \rangle_{\mathcal{H}_c}$



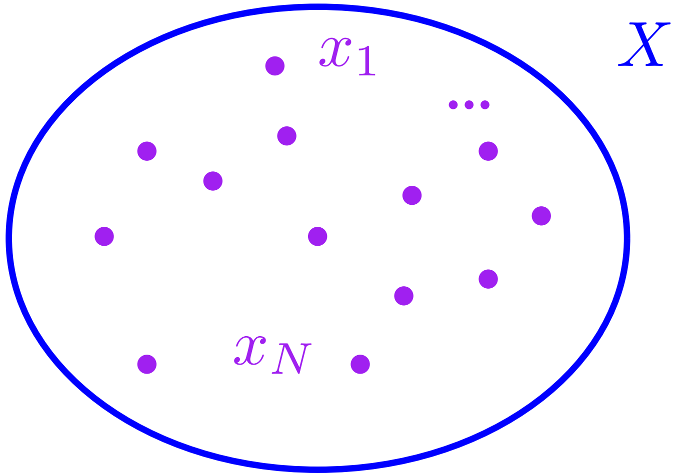
"Rotation" generated by $\frac{2}{\varepsilon}(V - V^*)$

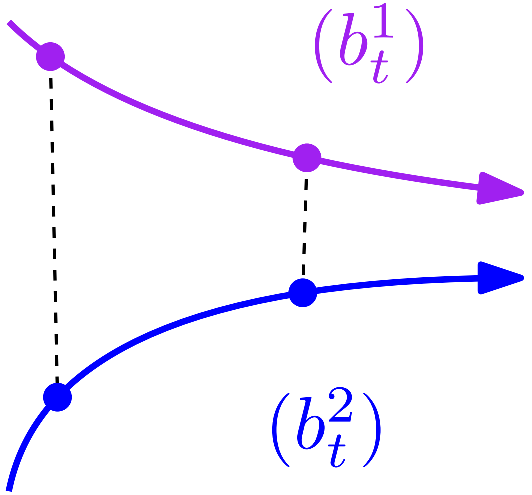
$\frac{2}{\varepsilon}(V - V^*)$ skew-symmetric: generates group of unitary operators, but unbounded.

Pressure: in the polar cone of $B(\mathcal{M}_+(X))$ for $\langle \cdot, \cdot \rangle_{\mathcal{H}_c}$, maintains $\mu_t \geq 0$.

Proof ideas of the main results

Proof idea for existence: approximate X by a finite space $X_N = \{x_1, \ldots, x_N\}$. For measures supported on X_N, the Sinkhorn flow is a maximal monotone evolution.





Proof idea of non-expansiveness: Monotone evolutions are non-expansive thus $\|b_t^1 - b_t^2\|_{\mathcal{H}_c}$ decreases.

Link with the Sinkhorn JKO scheme

Sinkhorn JKO:

$$\mu_{k+1}^\tau \in \arg\min_\mu \; E(\mu) + \frac{S_\varepsilon(\mu, \mu_k^\tau)}{2\tau}.$$

with S_ε Sinkhorn divergence and $E(\mu) = \int V \,\mathrm{d}\mu$ **potential energy**.

Sinkhorn potential flow

$$\frac{\varepsilon}{2}(\mathrm{Id} - K_{\mu_t}^2)^{-1} H_{\mu_t}[\dot{\mu}_t] + V + p_t = \text{Cst.}$$

Link with the Sinkhorn JKO scheme

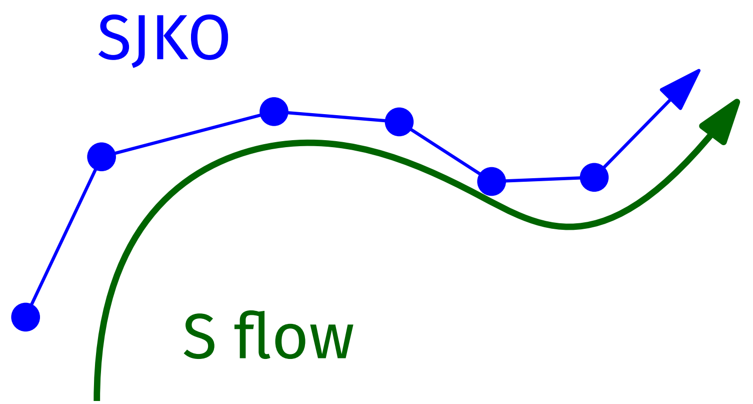
Sinkhorn JKO:

$$\mu_{k+1}^{\tau} \in \arg\min_{\mu} \; E(\mu) + \frac{S_\varepsilon(\mu, \mu_k^\tau)}{2\tau}.$$

with S_ε Sinkhorn divergence and $E(\mu) = \int V \,\mathrm{d}\mu$ **potential energy**.

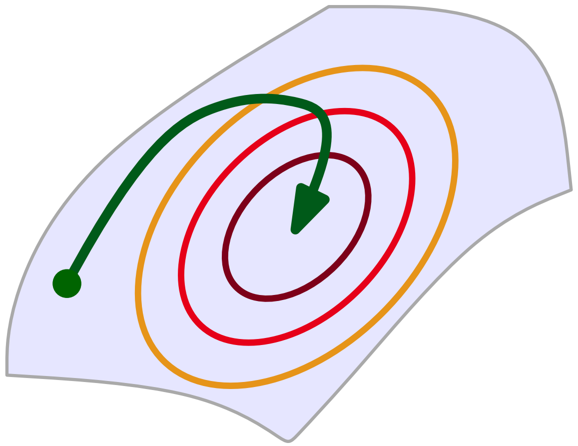
Sinkhorn potential flow

$$\frac{\varepsilon}{2}(\mathrm{Id} - K_{\mu_t}^2)^{-1} H_{\mu_t}[\dot{\mu}_t] + V + p_t = \text{Cst.}$$



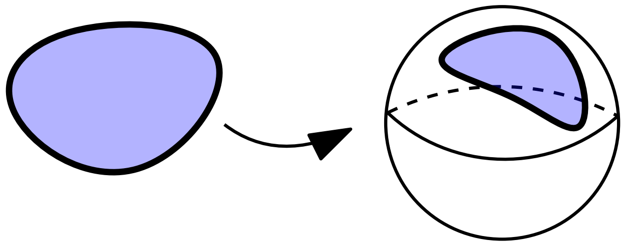
Proposition. If X is a **finite set**, the solutions of the Sinkhorn JKO scheme, properly interpolated in time, converge to the Sinkhorn flow as $\tau \to 0$ in $C([0,T], \mathcal{P}(X))$.

Future works



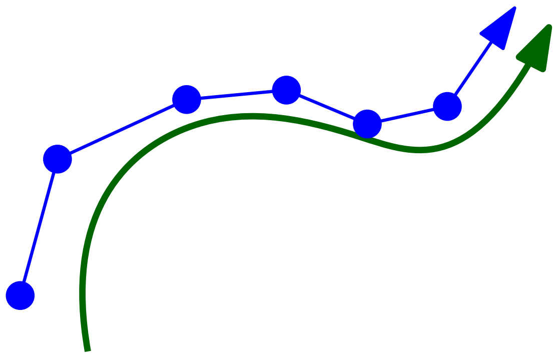
What I have not presented:

- our results on the geometry of Sinkhorn divergences.

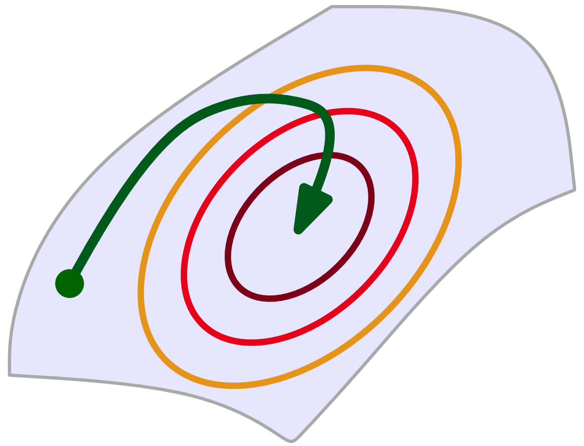


Some topics worth exploring:

- Going beyond potential energies,
- Limit $\varepsilon \to 0$ towards Wasserstein gradient flows,
- Use this flow for optimization or sampling?

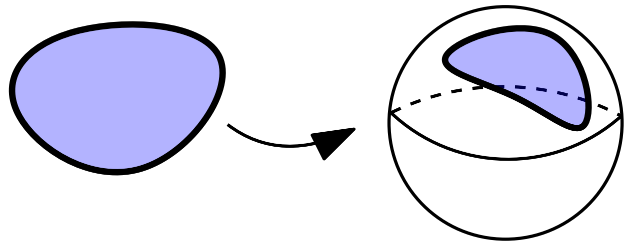


Future works



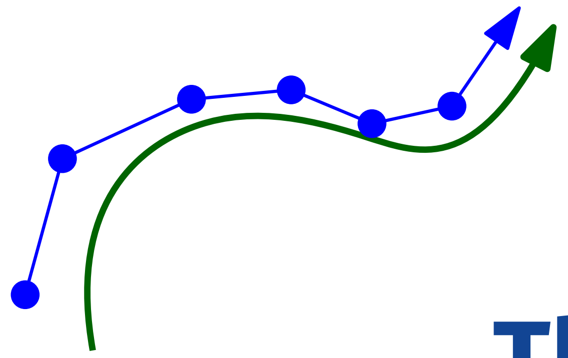
What I have not presented:

- our results on the geometry of Sinkhorn divergences.



Some topics worth exploring:

- Going beyond potential energies,
- Limit $\varepsilon \to 0$ towards Wasserstein gradient flows,
- Use this flow for optimization or sampling?



Thank you for your attention