Interaction-Force Transport Gradient Flows

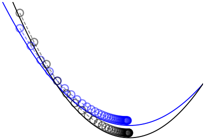
Egor Gladin (HU Berlin & HSE University), Pavel Dvurechensky (Weierstrass Institute),
Alexander Mielke (HU Berlin & Weierstrass Institute), **Jia-Jie Zhu** (Weierstrass Institute)

Weierstraß-Institut für
Angewandte Analysis und Stochastik

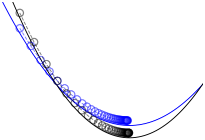
The Thirty-Eighth Annual Conference on Neural Information Processing Systems
(NeurIPS 2024)

From gradient descent to gradient flow

Optimization problem in $\mathbb{R}^d$: $\min_{x \in \mathbb{R}^d} F(x)$



From gradient descent to gradient flow

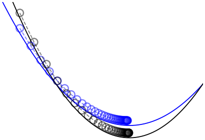


Optimization problem in $\mathbb{R}^d$: $\min_{x\in\mathbb{R}^d} F(x)$

Gradient descent $\quad x_{k+1} = x_k - \tau \cdot \nabla F(x_k)^\top$

Prox. step (implicit)/ JKO $\quad x_{k+1} \in \operatorname{argmin}_x \left(F(x) + \frac{1}{2\tau}\|x - x_k\|^2 \right)$

From gradient descent to gradient flow



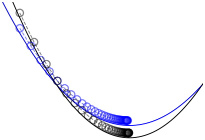
Optimization problem in $\mathbb{R}^d$: $\min_{x \in \mathbb{R}^d} F(x)$

Gradient descent $\quad x_{k+1} = x_k - \tau \cdot \nabla F(x_k)^\top$

Prox. step (implicit)/ JKO $\quad x_{k+1} \in \operatorname{argmin}_x \left(F(x) + \frac{1}{2\tau} \|x - x_k\|^2 \right)$

$\tau \to 0$ continuous time: ODE $\quad \dot{x}(t) = -\nabla F(x(t))^\top$

From gradient descent to gradient flow



Optimization problem in $\mathbb{R}^d$: $\min_{x \in \mathbb{R}^d} F(x)$

$$\text{Gradient descent} \quad x_{k+1} = x_k - \tau \cdot \nabla F(x_k)^\top$$

$$\text{Prox. step (implicit)/ JKO} \quad x_{k+1} \in \operatorname*{argmin}_x \left(F(x) + \frac{1}{2\tau} \|x - x_k\|^2 \right)$$

$$\tau \to 0 \text{ continuous time: ODE} \quad \dot{x}(t) = -\nabla F(x(t))^\top$$

is the *gradient-flow equation* of the energy $F(x)$ in the **space** $\mathbb{R}^d$ with the Eulidean geometry described by $\|x\|^2$.

From Euclidean gradient descent to Wasserstein gradient flow

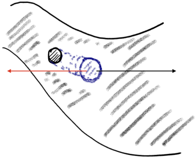
Generalizing the Euclidean geometry $\left(\mathbb{R}^d, \|\cdot\|\right)$ to $(\mathcal{P}, W_2)$

From Euclidean gradient descent to Wasserstein gradient flow

Generalizing the Euclidean geometry $\left(\mathbb{R}^d, \|\cdot\|\right)$ to $(\mathcal{P}, W_2)$

JKO: Wasserstein gradient flow [Otto, 1996, 2001]

$$\mu^{k+1} \in \underset{\mu \in \mathcal{P}}{\operatorname{argmin}}\, F(\mu) + \frac{1}{2\tau} {W_2}^2(\mu, \mu^k)$$



Continuous-time ($\tau \to 0$) **gradient flow equation**

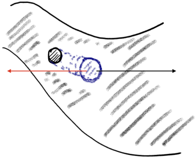
$$\partial_t \mu = -\mathrm{div}\left(\mu \nabla \frac{\delta F}{\delta \mu}[\mu]\right)$$

From Euclidean gradient descent to Wasserstein gradient flow

Generalizing the Euclidean geometry $\left(\mathbb{R}^d, \|\cdot\|\right)$ to $(\mathcal{P}, W_2)$

JKO: Wasserstein gradient flow [Otto, 1996, 2001]

$$\mu^{k+1} \in \operatorname*{argmin}_{\mu \in \mathcal{P}} F(\mu) + \frac{1}{2\tau} {W_2}^2(\mu, \mu^k)$$



Continuous-time ($\tau \to 0$) **gradient flow equation**

$$\partial_t \mu = -\mathrm{div}\left(\mu \nabla \frac{\delta F}{\delta \mu}[\mu]\right)$$

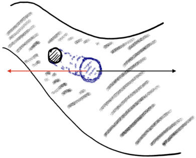
PDE has a **gradient structure**:
$$\begin{cases} \textbf{Measure Space}: & \mathcal{P} \text{ or } \mathcal{M}^+ \\ \text{Energy functional}: & F \text{ (e.g. KL)} \\ \text{Dissipation Geometry}: & W_2 \text{ or He} \end{cases}$$

From Euclidean gradient descent to Wasserstein gradient flow

Generalizing the Euclidean geometry $\left(\mathbb{R}^d, \|\cdot\|\right)$ to $(\mathcal{P}, W_2)$

JKO: Wasserstein gradient flow [Otto, 1996, 2001]

$$\mu^{k+1} \in \underset{\mu \in \mathcal{P}}{\text{argmin}}\, F(\mu) + \frac{1}{2\tau} W_2{}^2(\mu, \mu^k)$$



Continuous-time ($\tau \to 0$) **gradient flow equation**

$$\partial_t \mu = -\mathrm{div}\left(\mu \nabla \frac{\delta F}{\delta \mu}[\mu]\right)$$

PDE has a **gradient structure**:

- **Measure Space** : $\mathcal{P}$ or $\mathcal{M}^+$
- Energy functional : F (e.g. KL)
- Dissipation Geometry : W_2 or He

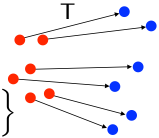
The merit of the right gradient flow formulation of a dissipative evolution equation is that it **separates energetics and kinetics**: The energetics endow the state space with a functional, the kinetics endow the **state space** with a (Riemannian) geometry via the metric tensor. [Otto 2001]

Wasserstein distance and optimal transport

"Euclidean distance" between probability measures

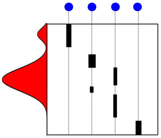
p-th order **Kantorovich-Wasserstein distance** between measures μ_0, μ_1 on $X \subset \mathbb{R}^d$ with p finite moments is defined through the Monge problem

$$W_p^p(\mu_0, \mu_1) := \min \left\{ \int |x - T(x)|^p \, \mathrm{d}\mu_0(x) \middle| \; T_\# \mu_0 = \mu_1 \right\}$$



the Kantorovich problem

$$W_p^p(\mu_0, \mu_1) := \min \left\{ \int |x_0 - x_1|^p \, \mathrm{d}\Pi \;\middle|\; \pi_\#^{(1)}\Pi = \mu_0, \pi_\#^{(2)}\Pi = \mu_1 \right\}$$



[Peyré and Cuturi, 2019]

Unbalanced transport: Hellinger-Kantorovich a.k.a. Wasserstein-Fisher-Rao

Best of both worlds

Wasserstein: transport/diffusion   Fisher-Rao / Hellinger: growth/birth-death

Unbalanced transport: Hellinger-Kantorovich a.k.a. Wasserstein-Fisher-Rao

Best of both worlds

Wasserstein: transport/diffusion   Fisher-Rao / Hellinger: growth/birth-death

H-K / W-FR [Chizat et al., 2018, 2019, Liero et al., 2018, Kondratyev et al., 2016] gradient flow:reaction-diffusion eqn.

Unbalanced transport: Hellinger-Kantorovich a.k.a. Wasserstein-Fisher-Rao

Best of both worlds

Wasserstein: transport/diffusion

Fisher-Rao / Hellinger: growth/birth-death

H-K / W-FR [Chizat et al., 2018, 2019, Liero et al., 2018, Kondratyev et al., 2016] gradient flow:reaction-diffusion eqn.

$$\dot{\mu} = -\alpha \mathbb{G}_W^{-1}(\mu) \frac{\delta F}{\delta \mu}[\mu]$$

Unbalanced transport: Hellinger-Kantorovich a.k.a. Wasserstein-Fisher-Rao

Best of both worlds

Wasserstein: transport/diffusion   Fisher-Rao / Hellinger: growth/birth-death

H-K / W-FR [Chizat et al., 2018, 2019, Liero et al., 2018, Kondratyev et al., 2016] gradient flow: reaction-diffusion eqn.

$$\dot{\mu} = -\alpha \mathbb{G}_W^{-1}(\mu) \frac{\delta F}{\delta \mu}[\mu] - \beta \mathbb{G}_{\mathrm{He}}^{-1}(\mu) \frac{\delta F}{\delta \mu}[\mu]$$

Unbalanced transport: Hellinger-Kantorovich a.k.a. Wasserstein-Fisher-Rao

Best of both worlds

Wasserstein: transport/diffusion Fisher-Rao / Hellinger: growth/birth-death

H-K / W-FR [Chizat et al., 2018, 2019, Liero et al., 2018, Kondratyev et al., 2016] gradient flow:reaction-diffusion eqn.

$$\dot{\mu} = -\alpha \mathbb{G}_W^{-1}(\mu) \frac{\delta F}{\delta \mu}[\mu] - \beta \mathbb{G}_{\mathrm{He}}^{-1}(\mu) \frac{\delta F}{\delta \mu}[\mu] = \alpha \mathrm{div}\left(\mu \nabla \frac{\delta F}{\delta \mu}\right) - \beta \mu \frac{\delta F}{\delta \mu}$$

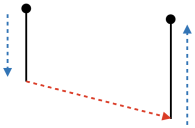
Unbalanced transport: Hellinger-Kantorovich a.k.a. Wasserstein-Fisher-Rao

Best of both worlds

Wasserstein: transport/diffusion   Fisher-Rao / Hellinger: growth/birth-death

H-K / W-FR [Chizat et al., 2018, 2019, Liero et al., 2018, Kondratyev et al., 2016] gradient flow:reaction-diffusion eqn.

$$\dot{\mu} = -\alpha \mathbb{G}_{W}^{-1}(\mu) \frac{\delta F}{\delta \mu}[\mu] - \beta \mathbb{G}_{\mathrm{He}}^{-1}(\mu) \frac{\delta F}{\delta \mu}[\mu] = \alpha \operatorname{div}\left(\mu \nabla \frac{\delta F}{\delta \mu}\right) - \beta \mu \frac{\delta F}{\delta \mu}$$



Kernelized UT: Interaction-force transport gradient flows

$$\min_{\mu\in\mathcal{P}}\left\{F(\mu) := \mathsf{MMD}^2(\mu,\pi)\right\} \quad \text{[Arbel et al., 2019] via WGF}$$

Kernelized UT: Interaction-force transport gradient flows

$$\min_{\mu\in\mathcal{P}}\left\{F(\mu) := \mathrm{MMD}^2(\mu,\pi)\right\} \quad \text{[Arbel et al., 2019] via WGF}$$

Application: generative model, inference, sampling

Kernelized UT: Interaction-force transport gradient flows

$$\min_{\mu \in \mathcal{P}} \left\{ F(\mu) := \mathsf{MMD}^2(\mu, \pi) \right\} \quad \text{[Arbel et al., 2019] via WGF}$$

Application: generative model, inference, sampling

[Gladin et al. & Z, 2024] Motivation: kernelize HK / WFR gradient flow

$$\dot{\mu} = -\alpha \cdot \mathbb{G}_W^{-1}(\mu) \frac{\delta F}{\delta \mu}[\mu] - \beta \cdot \mathbb{G}_{\mathsf{He}}^{-1}(\mu) \circ \mathcal{K}_\mu^{-1} \frac{\delta F}{\delta \mu}[\mu]$$

Kernelized UT: Interaction-force transport gradient flows

$$\min_{\mu\in\mathcal{P}}\left\{F(\mu) := \mathsf{MMD}^2(\mu,\pi)\right\} \quad \text{[Arbel et al., 2019] via WGF}$$

Application: generative model, inference, sampling

[Gladin et al. & Z, 2024] Motivation: kernelize HK / WFR gradient flow

$$\dot{\mu} = -\alpha\cdot\mathbb{G}_W^{-1}(\mu)\frac{\delta F}{\delta\mu}[\mu] - \beta\cdot\mathbb{G}_{\mathsf{He}}^{-1}(\mu)\circ\mathcal{K}_\mu^{-1}\frac{\delta F}{\delta\mu}[\mu]$$

$$\overset{\text{[ZM'24]}}{=} -\alpha\cdot\mathbb{G}_W^{-1}(\mu)\frac{\delta F}{\delta\mu}[\mu] - \beta\cdot\mathbb{G}_{\mathsf{MMD}}^{-1}\frac{\delta F}{\delta\mu}[\mu]$$

Kernelized UT: Interaction-force transport gradient flows

$$\min_{\mu\in\mathcal{P}}\left\{F(\mu) := \mathsf{MMD}^2(\mu,\pi)\right\} \quad \text{[Arbel et al., 2019] via WGF}$$

Application: generative model, inference, sampling

[Gladin et al. & Z, 2024] Motivation: kernelize HK / WFR gradient flow

$$\begin{aligned}\dot\mu &= -\alpha\cdot\mathbb{G}_W^{-1}(\mu)\frac{\delta F}{\delta\mu}[\mu] - \beta\cdot\mathbb{G}_{\mathsf{He}}^{-1}(\mu)\circ\mathcal{K}_\mu^{-1}\frac{\delta F}{\delta\mu}[\mu]\\ &\overset{\text{[ZM'24]}}{=} -\alpha\cdot\mathbb{G}_W^{-1}(\mu)\frac{\delta F}{\delta\mu}[\mu] - \beta\cdot\mathbb{G}_{\mathsf{MMD}}^{-1}\frac{\delta F}{\delta\mu}[\mu]\\ &= \alpha\cdot\mathrm{div}\left(\mu\int\nabla_2 k(x,\cdot)\,\mathrm{d}(\mu-\pi)(x)\right) - \beta\cdot(\mu-\pi)\end{aligned}$$

Kernelized UT: Interaction-force transport gradient flows

$$\min_{\mu \in \mathcal{P}} \left\{ F(\mu) := \mathsf{MMD}^2(\mu, \pi) \right\}$$ [Arbel et al., 2019] via WGF

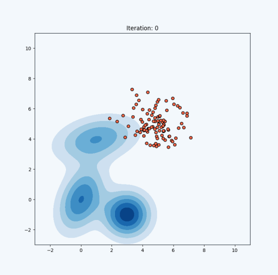
Application: generative model, inference, sampling

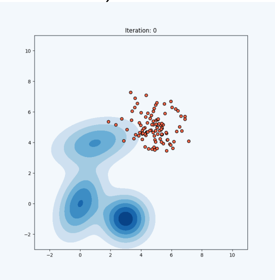
[Gladin et al. & Z, 2024] Motivation: kernelize HK / WFR gradient flow

$$\dot{\mu} = -\alpha \cdot \mathbb{G}_W^{-1}(\mu) \frac{\delta F}{\delta \mu}[\mu] - \beta \cdot \mathbb{G}_{\mathsf{He}}^{-1}(\mu) \circ \mathcal{K}_\mu^{-1} \frac{\delta F}{\delta \mu}[\mu]$$

$$\overset{\text{[ZM'24]}}{=} -\alpha \cdot \mathbb{G}_W^{-1}(\mu) \frac{\delta F}{\delta \mu}[\mu] - \beta \cdot \mathbb{G}_{\mathsf{MMD}}^{-1} \frac{\delta F}{\delta \mu}[\mu]$$

$$= \alpha \cdot \mathsf{div} \left(\mu \int \nabla_2 k(x, \cdot) \, \mathrm{d}(\mu - \pi)(x) \right) - \beta \cdot (\mu - \pi)$$





Kernelized UT: Interaction-force transport gradient flows

$$\min_{\mu\in\mathcal{P}}\left\{F(\mu):=\mathsf{MMD}^2(\mu,\pi)\right\}$$ [Arbel et al., 2019] via WGF

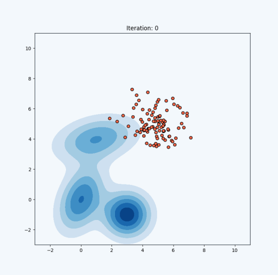
Application: generative model, inference, sampling

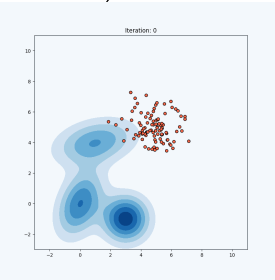
[Gladin et al. & Z, 2024] Motivation: kernelize HK / WFR gradient flow

$$\dot\mu=-\alpha\cdot\mathbb{G}_W^{-1}(\mu)\frac{\delta F}{\delta\mu}[\mu]-\beta\cdot\mathbb{G}_{\mathsf{He}}^{-1}(\mu)\circ\mathcal{K}_\mu^{-1}\frac{\delta F}{\delta\mu}[\mu]$$

$$\overset{\text{[ZM'24]}}{=}-\alpha\cdot\mathbb{G}_W^{-1}(\mu)\frac{\delta F}{\delta\mu}[\mu]-\beta\cdot\mathbb{G}_{\mathsf{MMD}}^{-1}\frac{\delta F}{\delta\mu}[\mu]$$

$$=\alpha\cdot\mathsf{div}\left(\mu\int\nabla_2 k(x,\cdot)\,\mathrm{d}(\mu-\pi)(x)\right)-\beta\cdot(\mu-\pi)$$





Kernelized UT: Interaction-force transport gradient flows

$$\min_{\mu\in\mathcal{P}}\left\{F(\mu) := \mathrm{MMD}^2(\mu,\pi)\right\} \quad \text{[Arbel et al., 2019] via WGF}$$

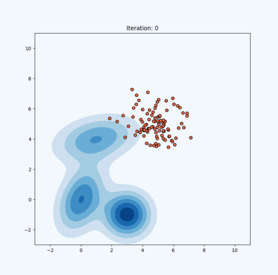
Application: generative model, inference, sampling

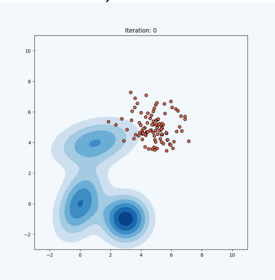
[Gladin et al. & Z, 2024] Motivation: kernelize HK / WFR gradient flow

$$\dot\mu = -\alpha\cdot\mathbb{G}_W^{-1}(\mu)\frac{\delta F}{\delta\mu}[\mu] - \beta\cdot\mathbb{G}_{\mathrm{He}}^{-1}(\mu)\circ\mathcal{K}_\mu^{-1}\frac{\delta F}{\delta\mu}[\mu]$$

$$\overset{\text{[ZM'24]}}{=} -\alpha\cdot\mathbb{G}_W^{-1}(\mu)\frac{\delta F}{\delta\mu}[\mu] - \beta\cdot\mathbb{G}_{\mathrm{MMD}}^{-1}\frac{\delta F}{\delta\mu}[\mu]$$

$$= \alpha\cdot\mathrm{div}\left(\mu\int\nabla_2 k(x,\cdot)\,\mathrm{d}(\mu-\pi)(x)\right) - \beta\cdot(\mu-\pi)$$





Analysis of interaction-force transport gradient flows

Theorem [Gladin et al. & Z, 2024] $\mathrm{MMD}^2(\mu, \pi)$ exponentially decays along the IFT gradient flow. In addition, KL decays under Bakry-Émery.

Analysis of interaction-force transport gradient flows

Theorem [Gladin et al. & Z, 2024] $\mathrm{MMD}^2(\mu, \pi)$ exponentially decays along the IFT gradient flow. In addition, KL decays under Bakry-Émery.

i.e., best of both worlds!

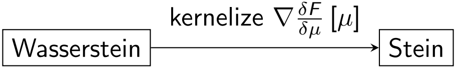
Interaction-Force Transport Gradient Flows

Egor Gladin (HU Berlin & HSE University), Pavel Dvurechensky (Weierstrass Institute),
Alexander Mielke (HU Berlin & Weierstrass Institute), **Jia-Jie Zhu** (Weierstrass Institute)

Thank you!

For code implementation, see: https://github.com/egorgladin/ift_flow

Gradient flow geometries obtained by kernelization



Theorem [Z and Mielke, 2024] The Riemannian metric tensors of Hellinger satisfy $\mathbb{G}_{\mathsf{MMD}} = \mathcal{K}_\mu \circ \mathbb{G}_{\mathsf{He}}(\mu)$, i.e.,

MMD=(de-)kernelized Hellinger