Nonlocal Wasserstein distance and the associated gradient flows

Dejan Slepčev
Carnegie Mellon University

Simons Institute
October 25, 2021.

Collaborations and References

- *Nonlocal Wasserstein distance* with Andrew Warren (in preparation)
- *Nonlocal-interaction equation on graphs: gradient flow structure and continuum limit* with Antonio Esposito, Francesco Patacchini, and André Schlichting, ARMA 2021
- *Clustering dynamics on graphs: from spectral clustering to mean shift through Fokker-Planck interpolation* with Katy Craig and Nicolás García Trillos

- *Gradient flows of the entropy for jump processes*, Erbar 2014, Annales IHP Prob and Stat.
- *Gromov–Hausdorff limit of Wasserstein spaces on point clouds* García Trillos 2020 Calc. Var. and PDE

Benamou–Brenier characterization on d_w

Consider probability measures with finite second moment $\rho_0, \rho_1 \in \mathcal{P}_2(\mathbb{R}^d)$
Let $\mathcal{A}$ be the set of paths from ρ_0 to ρ_1:

$$\mathcal{A} = \{(\rho, v) \ : \rho(\,\cdot\,, t) \in L^1(\mathbb{R}^d, [0, \infty)), v_t \in L^2(d\mu_t),$$
$$\boldsymbol{\partial_t \rho + \mathrm{div}(\rho\, v) = 0} \ \text{ on } \mathbb{R}^d \times [0, t]$$
$$\rho(\,\cdot\,, 0) = \rho_0 \text{ and } \rho(\,\cdot\,, 1) = \rho_1\}$$

Theorem

$$d_W^2(\rho_0, \rho_1) = \inf_{(\rho, v) \in \mathcal{A}} \int_0^1 \int \boldsymbol{|v(x, t)|^2 d\rho_t(x)}\, dt$$

Nonlocal calculus

Let $\eta : \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty]$ describe the nonlocal transportation connections. Assume $\eta(x, y) = \eta(y, x)$. Let $G = \{(x, y) \ : \ \eta(x, y) > 0\}$.

Nonlocal Gradient. Given $\varphi : \mathbb{R}^d \to \mathbb{R}$, the gradient $\overline{\nabla}\varphi : G \to \mathbb{R}$

$$\overline{\nabla}\varphi(x, y) = \varphi(y) - \varphi(x).$$

Nonlocal Divergence. Given $\boldsymbol{j} \in \mathcal{M}(G)$ – signed measures on G

$$\overline{\nabla} \cdot \boldsymbol{j}(x) = \int \eta(x, y)\boldsymbol{j}(x, dy).$$

For any $\boldsymbol{j} \in \mathcal{M}(G)$, its *nonlocal divergence* $\overline{\nabla} \cdot \boldsymbol{j} \in \mathcal{M}(\mathbb{R}^d)$ is defined as η-weighted adjoint of $\overline{\nabla}$, i.e.,

$$\int \phi \, d\overline{\nabla} \cdot \boldsymbol{j} = -\frac{1}{2} \iint_G \overline{\nabla}\phi(x, y)\eta(x, y) \, d\,\boldsymbol{j}(x, y).$$

Nonlocal continuity equation

What is the nonlocal analog of the continuity equation:

$$\partial_t \rho_t + \nabla \cdot \boldsymbol{j}_t = 0 \qquad \text{with flux} \qquad \boldsymbol{j}_t(x) = \rho_t(x) v_t(x) \ ?$$

Nonlocal fluxes j_t are defined on edges $(x, y) \in G$ while the densities are defined at points

$$\partial_t \rho_t(x) + (\overline{\nabla} \cdot \boldsymbol{j}_t)(x) = \partial_t \rho_t(x) + \int_G \eta(x, y) \boldsymbol{j}_t(x, dy) \ = 0 \ .$$

What is the nonlocal relationship between $\boldsymbol{j}$ and velocity v, given ρ_t?

Problem: There is no canonical way to define density along edges.
[Maas '11], [Mielke '11], [Chow, Huang, Li, Zhou '12], [Erbar '14]
use averaging functions $\theta : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$:

$$\boldsymbol{j}(x, y) = \theta\big(\rho(x), \rho(y)\big) \, v(x, y).$$

For any $r, s \geq 0$, $\theta(r, r) = r$, θ is increasing in r and s.

Nonlocal Wasserstein distance

Erbar '14, generalizes Maas '11

Given $\rho_0, \rho_1 \in \mathcal{P}_2(\mathbb{R}^d)$, consider vector fields $v : [0,1] \times G \to \mathbb{R}$ such that the solution of the continuity equation

$$\partial_t \rho(x) + \overline{\nabla} \cdot \big(\theta\big(\rho(x), \rho(y)\big)\, v(x,y)\big) = 0$$

with $\rho(0) = \rho_0$ and $\rho(1) = \rho_1$.

Admissible paths, $\mathrm{CE}(\rho_0, \rho_1)$, are all of the solutions $(\rho_t, v_t)_{t \in [0,1]}$ generated by above vector fields.

$$\mathcal{W}_\eta^2(\rho_0, \rho_1) := \frac{1}{2} \inf_{\mathrm{CE}(\rho_0,\rho_1)} \int_0^1 \int_G |v_t(x,y)|^2 |\theta\big(\rho_t(x), \rho_t(y)\big) \eta(x,y) dxdydt$$

On graphs

$$\mathcal{W}_\eta^2(\rho_0, \rho_1) := \frac{1}{2} \inf_{\mathrm{CE}(\rho_0,\rho_1)} \int_0^1 \sum_{x \in V} \sum_{y \in V} |v_t(x,y)|^2 |\theta\big(\rho_t(x), \rho_t(y)\big) \eta(x,y) dt$$

Interpolations considered

- $\theta(r, 0) = 0$, which includes geometric and logarithmic mean [Erbar, Maas]
- $\theta(r, 0) > 0$, which includes arithmetic mean. The set of tangent fluxes needs to be restricted to a cone.
- $\theta(r, s, j) = \begin{cases} r & \text{if } j > 0 \\ s & \text{if } j \leq 0 \end{cases}$ is the upwind interpolation. The resulting "distance" is not symmetric.

Graph Laplacian

- $V_n = \{x_1, \dots, x_n\}$, similarity matrix W:

$$W_{ij} := \eta\left(|x_i - x_j|\right).$$

  The weighted degree of a vertex is $d_i = \sum_j W_{i,j}$.

- Graph Laplacian

$$L = D - W,$$

  where $D = \mathrm{diag}(d_1, \dots, d_n)$.

- Graph heat equation

$$\frac{d}{dt}\rho(x_i) = -Lu(x_i).$$

Graph heat equation as graph Wasserstein gradient flows

[Maas '11], [Mielke '11], [Chow, Huang, Li, Zhou '12]
The graph heat equation

$$\frac{d}{dt}\rho = -L\rho$$

is the gradient of entropy

$$E(\rho) = \sum_{x \in V} \rho(x) \ln \rho(x)$$

with respect to the graph Wasserstein distance $d^2_{w,G}$ corresponding to

$$\theta(r, s) = \frac{r - s}{\ln r - \ln s} \quad \text{and } \theta(r, 0) = 0$$

For $F(\rho) = \sum_{x \in V} U(x)\rho(x)$ where U is a smooth function, and $\rho(\,\cdot\,, 0) = \delta_{x_1}$ the gradient flow is $\partial_t \rho(x) = 0$ for all x. The support of the solution cannot expand.

Graph heat equation as graph Wasserstein gradient flows

[Maas '11], [Mielke '11], [Chow, Huang, Li, Zhou '12]
The graph heat equation

$$\frac{d}{dt}\rho = -L\rho$$

is the gradient of entropy

$$E(\rho) = \sum_{x \in V} \rho(x) \ln \rho(x)$$

with respect to the graph Wasserstein distance $d^2_{w,G}$ corresponding to

$$\theta(r, s) = \frac{r - s}{\ln r - \ln s} \quad \text{and } \theta(r, 0) = 0$$

For $F(\rho) = \sum_{x \in V} U(x)\rho(x)$ where U is a smooth function, and $\rho(\,\cdot\,, 0) = \delta_{x_1}$ the gradient flow is $\partial_t \rho(x) = 0$ for all x. The support of the solution cannot expand.

Fractional Heat Equation

Nonlocal diffusion equation

$$\partial_t \rho = \int (\rho(y) - \rho(x))\eta(x-y)dy$$

is the gradient flow of entropy

$$E(\rho) = \int \rho(x) \ln \rho(x) dx$$

with respect to the gnonlocal Wasserstein distance $\mathcal{W}_\eta$ corresponding to

$$\theta(r,s) = \frac{r-s}{\ln r - \ln s} \quad \text{and } \theta(r,0) = 0$$

In particular, as Erbar observed, for $s \in (0,2)$ the fractional heat equation

$$\partial_t \rho = \int \frac{1}{|x-y|^{-d-s}} (\rho(y) - \rho(x))\, dy = \Delta^{s/2} \rho$$

is the gradient flow of entropy with respect to the nonlocal Wasserstein distance for $\eta(z) = \frac{1}{|z|^{-d-s}}$.

Rigorous definition of Nonlocal Wasserstein Distance

Difficulties:

- ρ may contain atoms
  $\Rightarrow$ measure valued framework
- Benamou-Brenier functional is not jointly convex in (ρ_t, v_t)
  $\Rightarrow$ flux variables
- Upwind metric is is only positively homogeneous:
  $g(v, v) \neq g(-v, -v)$. The geometric structure is Finslerian rather than Riemannian.
- In the general framework the underlying space ρ is supported within is described by measure μ. For most of the talk μ is the Lebesgue measure.

Nonlocal continuity equation and action

Nonlocal continuity equation in measure valued flux form

A pair $(\rho_t, \boldsymbol{j}_t)_{t\in[0,T]} \in \mathrm{CE}_T$ provided that $(\rho_t, \boldsymbol{j}_t) \in \mathcal{P}(\Omega) \times \mathcal{M}(G)$ for all $t \in [0, T]$:

$$\partial_t \rho_t + \overline{\nabla} \cdot \boldsymbol{j}_t = 0 \qquad \text{in } C_c^\infty([0,T) \times \Omega)^*$$

That is for smooth test functions φ

$$\int_0^T \int_\Omega \partial_t \varphi_t(x) d\rho_t(x) dt + \int_0^T \iint_G \overline{\nabla}\varphi_t(x,y)\, \eta(x,y) d\boldsymbol{j}_t(x,y) dt = 0 \,.$$

Nonlocal continuity equation and action

Nonlocal continuity equation in measure valued flux form

A pair $(\rho_t, \boldsymbol{j}_t)_{t\in[0,T]} \in \mathrm{CE}_T$ provided that $(\rho_t, \boldsymbol{j}_t) \in \mathcal{P}(\Omega) \times \mathcal{M}(G)$ for all $t \in [0, T]$:

$$\partial_t \rho_t + \overline{\nabla} \cdot \boldsymbol{j}_t = 0 \qquad \text{in } C_c^\infty([0,T) \times \Omega)^*$$

Action [for upwind flux]

For $\boldsymbol{j} \in \mathcal{M}(G)$, set $|\lambda| = |\rho \otimes \mu| + |\mu \otimes \rho| + |\boldsymbol{j}| \in \mathcal{M}^+(G)$ and define

$$\mathcal{A}(\mu; \rho, \boldsymbol{j}) = \iint_G \left(\alpha \left(\frac{d\boldsymbol{j}}{d|\lambda|}, \frac{d(\rho \otimes \mu)}{d|\lambda|} \right) + \alpha \left(-\frac{d\boldsymbol{j}}{d|\lambda|}, \frac{d(\mu \otimes \rho)}{d|\lambda|} \right) \right) \eta d|\lambda|.$$

where, the convex, and pos. one-homogeneous function α is defined by

$$\alpha(j, r) := \begin{cases} \frac{(j_+)^2}{r} & \text{if } r > 0, \qquad \text{with } j_+ = \max\{0, j\} \\ 0 & \text{if } j \leq 0 \text{ and } r = 0, \\ \infty & \text{if } j > 0 \text{ and } r = 0. \end{cases}$$

Finite action leads to upwind flux

Proposition

Let $(\rho, \boldsymbol{j}) \in \mathcal{P}(\Omega) \times \mathcal{M}(\Omega)$ such that $\mathcal{A}(\mu; \rho, \boldsymbol{j}) < \infty$, then:

- there exists a measurable nonlocal vector field $v : G \to \mathbb{R}$ such that

$$d\,\boldsymbol{j}(x, y). = v(x, y)_+ \, d\,\rho(x)\, d\,\mu(y) - v(x, y)_- \, d\,\mu(x)\, d\,\rho(y) \qquad \text{and}$$

$$\mathcal{A}(\mu; \rho, \boldsymbol{j}) = \iint_G \left(|v(x, y)_+|^2 + |v(y, x)_-|^2\right) \eta(x, y) d\rho(x) d\mu(y) \,.$$

- there exists an antisymmetric $\boldsymbol{j}^{as} \in \mathcal{M}^{as}(G)$ such that

$$\overline{\nabla}\cdot \boldsymbol{j} = \overline{\nabla}\cdot \boldsymbol{j}^{as}, \quad \text{that is} \quad \iint_G \overline{\nabla}\phi \, \eta \, d\,\boldsymbol{j} = \iint_G \overline{\nabla}\phi \, \eta \, d\,\boldsymbol{j}^{as} \quad \forall \phi \in C_c^\infty(\Omega),$$

  and an antisymmetric $v^{as} : G \to \mathbb{R}$ with

$$\mathcal{A}(\mu; \rho, \boldsymbol{j}^{as}) = 2 \iint_G |v^{as}(x, y)_+|^2 \eta d\rho(x) d\mu(y) \leq \mathcal{A}(\mu; \rho, \boldsymbol{j}).$$

Lower semicontinuity and compactness

Lower semicontinuity

if $\mu^n \rightharpoonup \mu$ in $\mathcal{M}(\Omega)$, $\rho^n \rightharpoonup \rho$ in $\mathcal{P}(\Omega)$, and $\boldsymbol{j}^n \rightharpoonup \boldsymbol{j}$ in $\mathcal{M}(G)$, then

$$\liminf_{n\to\infty} \mathcal{A}(\mu^n; \rho^n, \boldsymbol{j}^n) \geq \mathcal{A}(\mu; \rho, \boldsymbol{j})$$

Compactness

Let $(\rho^n, \boldsymbol{j}^n) \in \mathrm{CE}_T$ for each $n \in \mathbb{N}$ such that $\sup_n \int_0^T \mathcal{A}(\rho_t^n, \boldsymbol{j}_t^n)\, d\,t < \infty$. Then, there exists $(\rho, \boldsymbol{j}) \in \mathrm{CE}_T$ such that

$$\begin{aligned} \rho_t^n &\rightharpoonup \rho_t \quad \text{in } \mathcal{P}_2(\Omega) \text{ for all } t \in [0, T] \\ \boldsymbol{j}^n &\rightharpoonup \boldsymbol{j} \quad \text{in } \mathcal{M}_{\mathrm{loc}}(G \times [0, T]). \end{aligned}$$

Moreover $\liminf_{n\to\infty} \int_0^T \mathcal{A}(\rho_t^n, \boldsymbol{j}_t^n)\, d\,t \geq \int_0^T \mathcal{A}(\rho_t, \boldsymbol{j}_t)\, d\,t.$

Compactness of solutions to CE

Assumption (weight function)

The μ-measurable nonnegative symmetric lsc. function $\eta\colon G \to \mathbb{R}$ satisfies:

- The measure $\eta(\cdot,\cdot)d\mu$ is uniformly integrable close to diagonal, that is

$$\lim_{\varepsilon\to 0} \sup_{x\in\Omega} \int_{B_\varepsilon(x)} |x-y|^2 \eta(x,y) d\mu(y) = 0, \quad B_\varepsilon(x) = \{y \in \Omega : |x-y| < \varepsilon\}.$$

Compactness: Let $(\rho^n, \boldsymbol{j}^n) \in \mathrm{CE}_T$ for each $n \in \mathbb{N}$ such that $\sup_{n\in\mathbb{N}} M_2(\rho_0^n) < \infty$ and $\sup_n \int_0^T \mathcal{A}(\rho_t^n, \boldsymbol{j}_t^n)\, d\,t < \infty$. Then, there exists $(\rho, \boldsymbol{j}) \in \mathrm{CE}_T$ such that

$$\begin{aligned} \rho_t^n &\rightharpoonup \rho_t \quad \text{in } \mathcal{P}_2(\Omega) \text{ for all } t \in [0,T] \\ \boldsymbol{j}^n &\rightharpoonup \boldsymbol{j} \quad \text{in } \mathcal{M}_{\mathrm{loc}}(G \times [0,T]). \end{aligned}$$

Moreover, the action is lower semicontinuous

$$\liminf_{n\to\infty} \int_0^T \mathcal{A}(\rho_t^n, \boldsymbol{j}_t^n)\, d\,t \geq \int_0^T \mathcal{A}(\rho_t, \boldsymbol{j}_t)\, d\,t.$$

Upwind nonlocal Wasserstein distance

Definition

For $\rho_0, \rho_1 \in \mathcal{P}_2(\Omega)$ the *nonlocal upwind Wasserstein quasimetric*

$$\mathcal{W}_\eta(\rho_0, \rho_1)^2 = \inf \left\{ \int_0^1 \mathcal{A}(\rho_t, \boldsymbol{j}_t) dt \ : \ (\rho, \boldsymbol{j}) \in \mathrm{CE}(\rho_0, \rho_1) \right\} .$$

Properties:

- Minimum is attained for $(\rho, \boldsymbol{j}) \in \mathrm{CE}(\rho_0, \rho_1)$ with $\mathcal{A}(\rho_t, \boldsymbol{j}_t) = \mathcal{W}_\eta(\rho_0, \rho_1)^2$.
- $\mathcal{W}_\eta$ is jointly narrowly lower semicontinuous.
- For upwind interpolation we will use $\mathcal{T}_\eta$ instead of $\mathcal{W}_\eta$. Note that $\mathcal{T}\eta$ is not symmetric.

Expell cost

Proposition [Warren and S.]

- If η is nonintegrable: $\eta(x,y) = \boldsymbol{\eta}(|x-y|)$ and $\boldsymbol{\eta}(r) > r^{-d-s}$ when $r \leq \delta$, then
$$\mathcal{W}_\eta(\delta_0, c\chi_{B(0,\delta)}) \lesssim \delta^{s/2}.$$
- If $\eta(x,y) = \boldsymbol{\eta}(x-y)$ and $\int_{\mathbb{R}^d} \boldsymbol{\eta} dx < \infty$ and we consider arithmetic mean or upwind interpolation then there exists $c > 0$ such that for all $\nu \perp \rho$
$$\mathcal{W}_\eta(\delta_0, \nu) \geq c.$$
- If $\eta(x,y) = \boldsymbol{\eta}(x-y)$ and $\int_{\mathbb{R}^d} \boldsymbol{\eta} dx < \infty$ and we consider interpolation with $\theta(r,0) = 0$ then $\nu \perp \rho$
$$\mathcal{W}_\eta(\delta_0, \nu) = \infty.$$

Furthermore $\mathcal{T}_\eta(\nu, \delta_0) = \infty$.

Topology

Erbar showed that the topology metrized by NLW is at least as strong as the the one generated by Wasserstein metric (narrow convergence plus moment control)

Proposition [Warren and S.]

- If η is nonintegrable: $\eta(x, y) = \boldsymbol{\eta}(|x - y|)$ and $\boldsymbol{\eta}(r) > r^{-d-s}$ when $r \leq \delta$, then the topology generated by $\mathcal{W}_\eta$ on $\mathcal{P}((B(0, R))$ is the weak topology.
- If η is integrable: If $\eta(x, y) = \boldsymbol{\eta}(x - y)$ and $\int_{\mathbb{R}^d} \boldsymbol{\eta} dx < \infty$ and we consider arithmetic mean then there exist $0 < c_1, c_2$ such that for all $\rho_0, \rho_1 \in \mathcal{P}((B(0, R))$

$$c_1 \|\rho_1 - \rho_0\|_{TV} \leq \mathcal{W}(\rho_0, \rho_1) \leq c_2 \left(\|\rho_1 - \rho_0\|_{TV} + R\sqrt{d_{Monge}(\rho_0, \rho_1)} \right)$$

Error bounds

Theorem [Warren and S.]

Upper bound. In the cases where the expel cost is finite (and not including upwind interpolation) and η compactly supported

$$\varepsilon \mathcal{W}_{\varepsilon,\eta}(\mu_0, \mu_1) \leq \left(\frac{1}{\sqrt{\sigma_\eta}} \right) d_W(\mu_0, \mu_1) + O(\sqrt{\varepsilon}).$$

where $\eta_\varepsilon(z) = \frac{1}{\varepsilon^d} \eta \frac{z}{\varepsilon}$, ω_d is the volume of the unit ball, and

$$\sigma_\eta = \frac{1}{d} \int_{\mathbb{R}^d} |x|^2 \eta(x) dx.$$

Lower bound. Assume both $\mu_0, \mu_1 \in \mathcal{P}(\mathbb{R}^d)$ are supported inside $\overline{B}(0, R)$ with $R \geq 1$ and η compactly supported. Then,

$$d_W^2(\mu_0, \mu_1) \leq \varepsilon^2 \sigma_\eta \mathcal{W}_{\eta,\varepsilon}^2(\mu_0, \mu_1) + CR^2 \sqrt{\varepsilon}.$$

Upper bound: elements of the proof I

We use the Wasserstein geodesic to build a competitor. This includes two levels of smoothing.

1. Exact solutions to nonlocal transport. Let

$$\zeta(r) = \int_r^\infty s\eta(s)ds.$$

Consider a solution of the continuity equation

$$\partial_t \rho + \nabla \cdot J = 0.$$

Let $\rho_\zeta = \rho * \zeta$ and $J_\zeta = J * \zeta$. Then $\partial_t \rho_\zeta + \nabla \cdot (J_\zeta) = 0$.

Let $j(x,y) = (y-x) \cdot J(y)$. Then

$$\partial_t \rho_\zeta + \overline{\nabla} \cdot j = \partial_t \rho_\zeta + \int j(x,y)\eta(x-y)dy = 0.$$

Upper bound: elements of the proof II

We use the Wasserstein geodesic to build a competitor. This includes two levels of smoothing.

2. Smoothing that controls the interpolation. The problem arises if $\rho(x) - \theta(\rho(x), \rho(y)) > c > 0$.
Let $K(x) = c\, e^{-|x|}$ where $\frac{1}{c} = \int_{\mathbb{R}^d} e^{-|x|} dx$. Let $K_\delta(x) = \frac{1}{\delta^d} K\left(\frac{x}{\delta}\right)$.

Consider $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. Let $\mu_\delta = \mu * K_\delta$
If $|y - x| < \delta$,

$$\mu_\delta(y) \leq \mu_\delta(x)\left(1 + \frac{3}{\delta}|y - x|\right).$$

Lower bound: elements of the proof I

We use the dual formulation to provide competitor. In particular the nonlocal Hamilton-Jacobi equation. For graphs developed in *Gangbo, Li, Mou '19*, *Erbar, Maas, Wirth '20*.

Background:
Lemma. Suppose that μ_0 and μ_1 are probability measures supported within $B(0, R)$. Then,

$$\frac{1}{2} d_W^2(\mu_0, \mu_1) = \sup_{\phi_t \in BL([0,1]\times\mathbb{R}^d)} \left\{ \int \phi_1 d\mu_1 - \int \phi_0 d\mu_0 : \partial_t \phi_t + \frac{1}{2} |\nabla \phi_t|^2 \leq 0 \right\}$$

it holds that the optimal Hamilton-Jacobi subsolution has the property that $\mathrm{Lip}(\phi_t) \leq 2R$, for Lebesgue-almost all $t \in [0, 1]$.

Lower bound: elements of the proof II

A Lipschitz function $\phi : \mathbb{R}^d \times [0,1] \to \mathbb{R}$ is a nonlocal Hamilton-Jacobi subsolution, $\phi_t \in \mathrm{HJ}^1_{\mathrm{NL}}$ if, for a.e. $t \in (0,T)$ for all probability measures $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, and for any λ such that $Leb \ll \lambda$,

$$\int \partial_t \phi_t d\mu + \frac{1}{2} \int (\phi_t(y) - \phi_t(x))^2 \theta \left(\frac{d\mu}{d\lambda}(x), \frac{d\mu}{d\lambda}(y) \right) \eta_\varepsilon(x,y) d\lambda(x) d\lambda(y) \leq 0.$$

Then, the duality formula we expect to hold is

$$\frac{1}{2} \mathcal{W}^2_{\eta,\varepsilon}(\mu_0, \mu_1) = \sup \left\{ \int \phi_1(x) d\mu_1(x) - \int \phi_0(x) \mu_0(x) : \phi_t \in \mathrm{HJ}^1_{\mathrm{NL}} \right\}.$$

For technical reasons, we introduce a “smoothed version” of the nonlocal Wasserstein distance instead.