

A Scalable Statistical Theory for Smooth Wasserstein Distances

Ziv Goldfeld

Cornell University

February 9th, 2021

Statistical Divergences

Statistical Divergences

Definition: Discrepancy measure $\delta : \mathcal{P}(\mathbb{R}^d)^2 \rightarrow [0, \infty]$ btw probability distributions s.t. $\delta(\mu, \nu) = 0 \iff \mu = \nu$

Statistical Divergences

Definition: Discrepancy measure $\delta : \mathcal{P}(\mathbb{R}^d)^2 \rightarrow [0, \infty]$ btw probability distributions s.t. $\delta(\mu, \nu) = 0 \iff \mu = \nu$

Inference/learning tasks: Optimize/threshold/evaluate a divergence

Statistical Divergences

Definition: Discrepancy measure $\delta : \mathcal{P}(\mathbb{R}^d)^2 \rightarrow [0, \infty]$ btw probability distributions s.t. $\delta(\mu, \nu) = 0 \iff \mu = \nu$

Inference/learning tasks: Optimize/threshold/evaluate a divergence

Example: Implicit generative modeling to learn $\nu_{\theta} \approx \mu \in \mathcal{P}(\mathbb{R}^d)$

Statistical Divergences

Definition: Discrepancy measure $\delta : \mathcal{P}(\mathbb{R}^d)^2 \rightarrow [0, \infty]$ btw probability distributions s.t. $\delta(\mu, \nu) = 0 \iff \mu = \nu$

Inference/learning tasks: Optimize/threshold/evaluate a divergence

Example: Implicit generative modeling to learn $\nu_{\theta} \approx \mu \in \mathcal{P}(\mathbb{R}^d)$

- Map $Z \sim \nu_Z \in \mathcal{P}(\mathbb{R}^p)$, $p \ll d$, to \mathbb{R}^d space via transformation $\nu_{X|Z}^{(\theta)}$

Statistical Divergences

Definition: Discrepancy measure $\delta : \mathcal{P}(\mathbb{R}^d)^2 \rightarrow [0, \infty]$ btw probability distributions s.t. $\delta(\mu, \nu) = 0 \iff \mu = \nu$

Inference/learning tasks: Optimize/threshold/evaluate a divergence

Example: Implicit generative modeling to learn $\nu_\theta \approx \mu \in \mathcal{P}(\mathbb{R}^d)$

- Map $Z \sim \nu_Z \in \mathcal{P}(\mathbb{R}^p)$, $p \ll d$, to \mathbb{R}^d space via transformation $\nu_{X|Z}^{(\theta)}$
- ⇒ **Generative model:** $\nu_\theta(\cdot) := \int_{\mathbb{R}^p} \nu_{X|Z}^{(\theta)}(\cdot|z) d\nu_Z(z)$

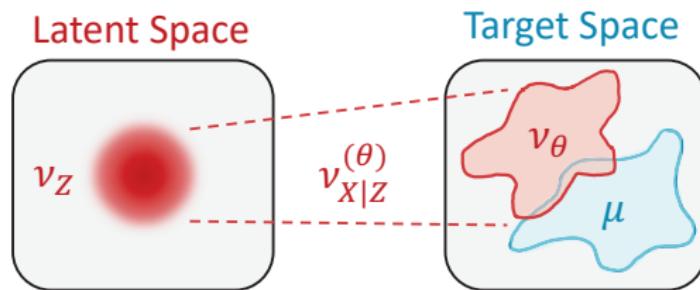
Statistical Divergences

Definition: Discrepancy measure $\delta : \mathcal{P}(\mathbb{R}^d)^2 \rightarrow [0, \infty]$ btw probability distributions s.t. $\delta(\mu, \nu) = 0 \iff \mu = \nu$

Inference/learning tasks: Optimize/threshold/evaluate a divergence

Example: Implicit generative modeling to learn $\nu_\theta \approx \mu \in \mathcal{P}(\mathbb{R}^d)$

- Map $Z \sim \nu_Z \in \mathcal{P}(\mathbb{R}^p)$, $p \ll d$, to \mathbb{R}^d space via transformation $\nu_{X|Z}^{(\theta)}$
- ⇒ **Generative model:** $\nu_\theta(\cdot) := \int_{\mathbb{R}^p} \nu_{X|Z}^{(\theta)}(\cdot|z) d\nu_Z(z)$



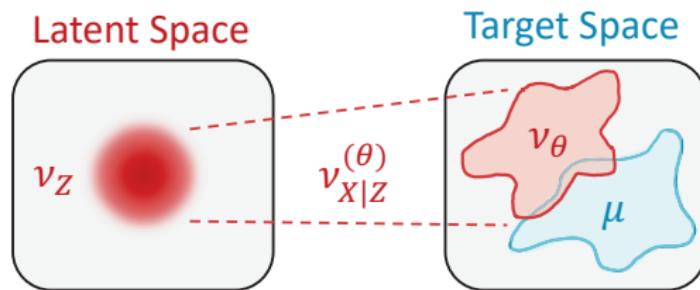
Statistical Divergences

Definition: Discrepancy measure $\delta : \mathcal{P}(\mathbb{R}^d)^2 \rightarrow [0, \infty]$ btw probability distributions s.t. $\delta(\mu, \nu) = 0 \iff \mu = \nu$

Inference/learning tasks: Optimize/threshold/evaluate a divergence

Example: Implicit generative modeling to learn $\nu_\theta \approx \mu \in \mathcal{P}(\mathbb{R}^d)$

- Map $Z \sim \nu_Z \in \mathcal{P}(\mathbb{R}^p)$, $p \ll d$, to \mathbb{R}^d space via transformation $\nu_{X|Z}^{(\theta)}$
- ⇒ **Generative model:** $\nu_\theta(\cdot) := \int_{\mathbb{R}^p} \nu_{X|Z}^{(\theta)}(\cdot|z) d\nu_Z(z)$



Minimum distance estimation: Solve

$$\theta^\star \in \operatorname{argmin}_\theta \delta(\mu, \nu_\theta)$$

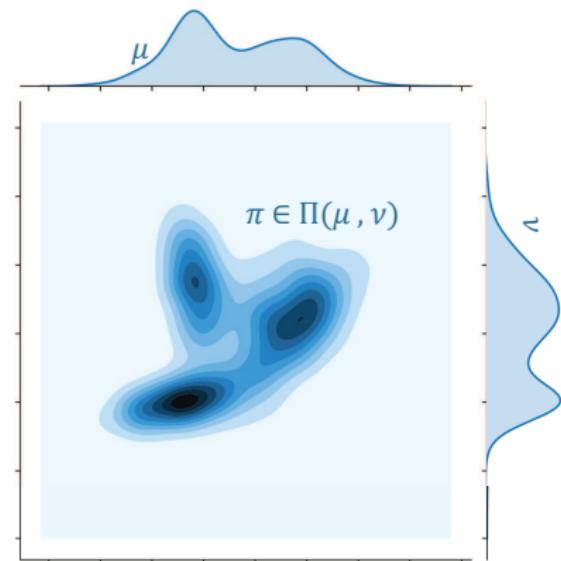
The p -Wasserstein Distance

Setup: $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite p th moments)

The p -Wasserstein Distance

Setup: $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite p th moments)

- **Coupling:** $\Pi(\mu, \nu) = \left\{ \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \pi(\cdot \times \mathbb{R}^d) = \mu, \pi(\mathbb{R}^d \times \cdot) = \nu \right\}$



The p -Wasserstein Distance

Setup: $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite p th moments)

- **Coupling:** $\Pi(\mu, \nu) = \left\{ \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \pi(\cdot \times \mathbb{R}^d) = \mu, \pi(\mathbb{R}^d \times \cdot) = \nu \right\}$
- **Cost:** $c(x, y) = \|x - y\|^p$ for transporting x to y

The p -Wasserstein Distance

Setup: $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite p th moments)

- **Coupling:** $\Pi(\mu, \nu) = \left\{ \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \pi(\cdot \times \mathbb{R}^d) = \mu, \pi(\mathbb{R}^d \times \cdot) = \nu \right\}$
- **Cost:** $c(x, y) = \|x - y\|^p$ for transporting x to y

Definition (p -Wasserstein distance)

For $p \geq 1$ define: $W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p d\pi(x, y) \right)^{\frac{1}{p}}$

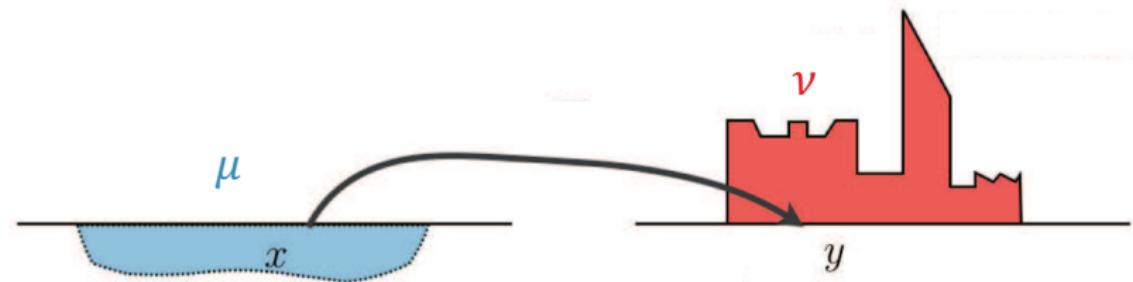
The p -Wasserstein Distance

Setup: $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite p th moments)

- **Coupling:** $\Pi(\mu, \nu) = \left\{ \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \pi(\cdot \times \mathbb{R}^d) = \mu, \pi(\mathbb{R}^d \times \cdot) = \nu \right\}$
- **Cost:** $c(x, y) = \|x - y\|^p$ for transporting x to y

Definition (p -Wasserstein distance)

For $p \geq 1$ define: $W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p d\pi(x, y) \right)^{\frac{1}{p}}$



The p -Wasserstein Distance

Setup: $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite p th moments)

- **Coupling:** $\Pi(\mu, \nu) = \left\{ \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \pi(\cdot \times \mathbb{R}^d) = \mu, \pi(\mathbb{R}^d \times \cdot) = \nu \right\}$
- **Cost:** $c(x, y) = \|x - y\|^p$ for transporting x to y

Definition (p -Wasserstein distance)

For $p \geq 1$ define: $W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p d\pi(x, y) \right)^{\frac{1}{p}}$

Comments:

The p -Wasserstein Distance

Setup: $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite p th moments)

- **Coupling:** $\Pi(\mu, \nu) = \left\{ \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \pi(\cdot \times \mathbb{R}^d) = \mu, \pi(\mathbb{R}^d \times \cdot) = \nu \right\}$
- **Cost:** $c(x, y) = \|x - y\|^p$ for transporting x to y

Definition (p -Wasserstein distance)

For $p \geq 1$ define: $W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p d\pi(x, y) \right)^{\frac{1}{p}}$

Comments:

- **Robust to support mismatch:** $W_p(\mu, \nu) < \infty, \forall \mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$

The p -Wasserstein Distance

Setup: $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite p th moments)

- **Coupling:** $\Pi(\mu, \nu) = \left\{ \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \pi(\cdot \times \mathbb{R}^d) = \mu, \pi(\mathbb{R}^d \times \cdot) = \nu \right\}$
- **Cost:** $c(x, y) = \|x - y\|^p$ for transporting x to y

Definition (p -Wasserstein distance)

For $p \geq 1$ define: $W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p d\pi(x, y) \right)^{\frac{1}{p}}$

Comments:

- **Robust to support mismatch:** $W_p(\mu, \nu) < \infty, \forall \mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$
- **Metric:** $(\mathcal{P}_p(\mathbb{R}^d), W_p)$ is metric space (metrizes weak convergence)

The p -Wasserstein Distance

Setup: $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite p th moments)

- **Coupling:** $\Pi(\mu, \nu) = \left\{ \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \pi(\cdot \times \mathbb{R}^d) = \mu, \pi(\mathbb{R}^d \times \cdot) = \nu \right\}$
- **Cost:** $c(x, y) = \|x - y\|^p$ for transporting x to y

Definition (p -Wasserstein distance)

For $p \geq 1$ define: $W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p d\pi(x, y) \right)^{\frac{1}{p}}$

Comments:

- **Robust to support mismatch:** $W_p(\mu, \nu) < \infty, \forall \mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$
- **Metric:** $(\mathcal{P}_p(\mathbb{R}^d), W_p)$ is metric space (metrizes weak convergence)
- **Duality:** $W_1(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \int f d(\mu - \nu)$

The p -Wasserstein Distance

Setup: $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite p th moments)

- **Coupling:** $\Pi(\mu, \nu) = \left\{ \pi \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) : \pi(\cdot \times \mathbb{R}^d) = \mu, \pi(\mathbb{R}^d \times \cdot) = \nu \right\}$
- **Cost:** $c(x, y) = \|x - y\|^p$ for transporting x to y

Definition (p -Wasserstein distance)

For $p \geq 1$ define: $W_p(\mu, \nu) := \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^p d\pi(x, y) \right)^{\frac{1}{p}}$

Comments:

- **Robust to support mismatch:** $W_p(\mu, \nu) < \infty, \forall \mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$
- **Metric:** $(\mathcal{P}_p(\mathbb{R}^d), W_p)$ is metric space (metrizes weak convergence)
- **Duality:** $W_1(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \int f d(\mu - \nu) \xrightarrow{\text{MDE}} \text{W-GAN}$ [Arjovsky et al'17]

From Duality to Generative Adversarial Networks

Dual representation: $W_1(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \mathbb{E}_\mu[f(X)] - \mathbb{E}_\nu[f(Y)]$

From Duality to Generative Adversarial Networks

Dual representation: $W_1(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \mathbb{E}_\mu[f(X)] - \mathbb{E}_\nu[f(Y)]$

W-GAN [Arjovsky et al'17]:

From Duality to Generative Adversarial Networks

Dual representation: $W_1(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \mathbb{E}_{\mu}[f(\mathbf{X})] - \mathbb{E}_{\nu}[f(Y)]$

W-GAN [Arjovsky et al'17]:

- μ (\mathbf{X} data sample)

From Duality to Generative Adversarial Networks

Dual representation: $W_1(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \mathbb{E}_{\mu}[f(\mathbf{X})] - \mathbb{E}_{\nu}[f(\mathbf{Y})]$

W-GAN [Arjovsky et al'17]:

- μ (\mathbf{X} data sample)
- $\nu = \nu_{\theta}$ ($\mathbf{Y} = g_{\theta}(Z)$ gen. sample)

From Duality to Generative Adversarial Networks

Dual representation: $W_1(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \mathbb{E}_{\mu}[f(X)] - \mathbb{E}_{\nu}[f(Y)]$

W-GAN [Arjovsky et al'17]:

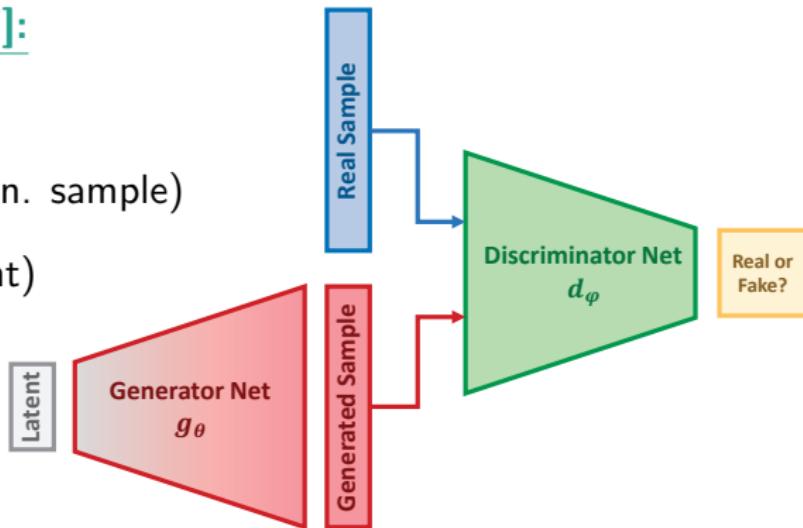
- μ (X data sample)
- $\nu = \nu_{\theta}$ ($Y = g_{\theta}(Z)$ gen. sample)
- $f = d_{\phi}$ (Lip_1 constraint)

From Duality to Generative Adversarial Networks

Dual representation: $W_1(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \mathbb{E}_{\mu}[f(X)] - \mathbb{E}_{\nu}[f(Y)]$

W-GAN [Arjovsky et al'17]:

- μ (X data sample)
- $\nu = \nu_{\theta}$ ($Y = g_{\theta}(Z)$ gen. sample)
- $f = d_{\phi}$ (Lip_1 constraint)

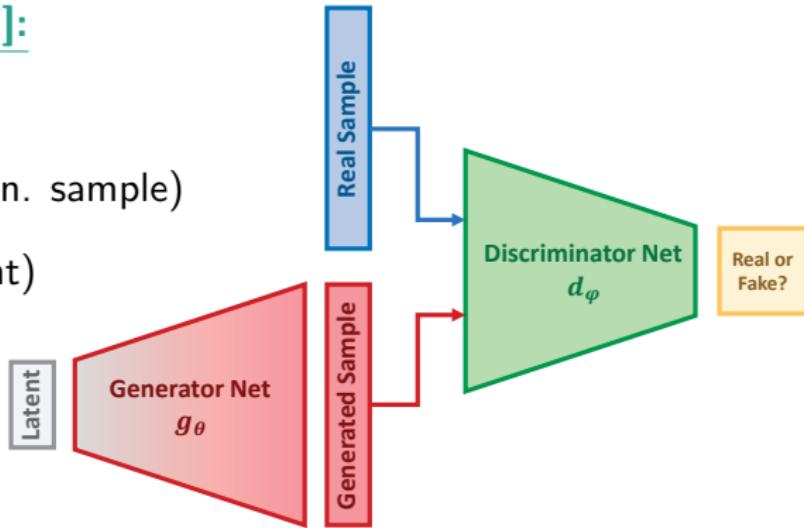


From Duality to Generative Adversarial Networks

Dual representation: $W_1(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \mathbb{E}_{\mu}[f(X)] - \mathbb{E}_{\nu}[f(Y)]$

W-GAN [Arjovsky et al'17]:

- μ (X data sample)
- $\nu = \nu_{\theta}$ ($Y = g_{\theta}(Z)$ gen. sample)
- $f = d_{\phi}$ (Lip_1 constraint)



$$\implies \inf_{\theta} W_1(\mu, \nu_{\theta}) \cong \inf_{\theta} \sup_{\varphi: d_{\varphi} \in \text{Lip}_1(\mathbb{R}^d)} \mathbb{E}[d_{\varphi}(X)] - \mathbb{E}[d_{\varphi}(g_{\theta}(Z))]$$

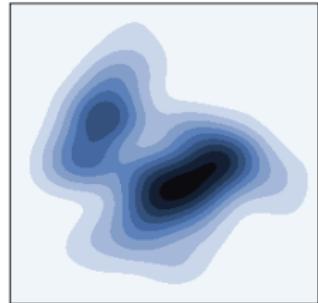
Generative Adversarial Networks

NVIDIA's ProGAN 2.0 [Karras *et al*'19]



Empirical Approximation in High Dimensions

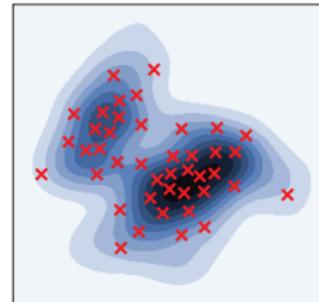
In practice: We don't have $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ but samples



Empirical Approximation in High Dimensions

In practice: We don't have $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ but samples

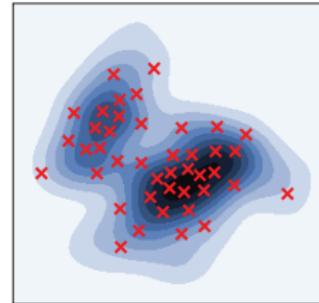
- **Samples:** $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mu$ and $Y_1, \dots, Y_n \stackrel{\text{i.i.d.}}{\sim} \nu$



Empirical Approximation in High Dimensions

In practice: We don't have $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ but samples

- **Samples:** $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mu$ and $Y_1, \dots, Y_n \stackrel{\text{i.i.d.}}{\sim} \nu$
- **Empirical dist.:** $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ and $\nu_n := \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$

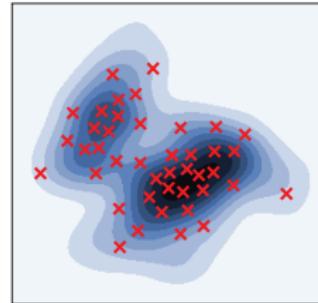


Empirical Approximation in High Dimensions

In practice: We don't have $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ but samples

- **Samples:** $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mu$ and $Y_1, \dots, Y_n \stackrel{\text{i.i.d.}}{\sim} \nu$
- **Empirical dist.:** $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ and $\nu_n := \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$

⇒ Can we approximate $W_p(\mu, \nu) \approx W_p(\mu_n, \nu_n)$?

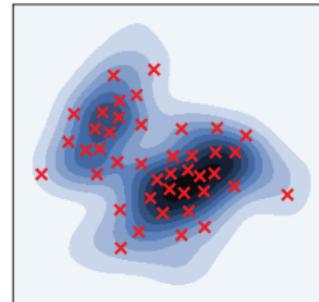


Empirical Approximation in High Dimensions

In practice: We don't have $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ but samples

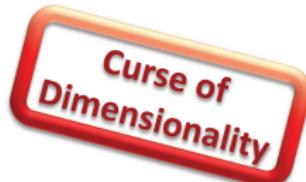
- **Samples:** $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mu$ and $Y_1, \dots, Y_n \stackrel{\text{i.i.d.}}{\sim} \nu$
- **Empirical dist.:** $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ and $\nu_n := \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$

⇒ Can we approximate $W_p(\mu, \nu) \approx W_p(\mu_n, \nu_n)$?



Theorem (Dudley'69, Boissard-Le Gouic'14, Niles Weed-Bach'19,..)

For $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ and $d > 2p$: $\mathbb{E}[|W_p(\mu, \nu) - W_p(\mu_n, \nu_n)|] \asymp n^{-\frac{1}{d}}$

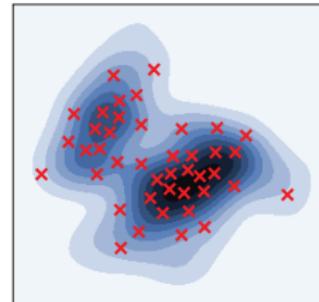


Empirical Approximation in High Dimensions

In practice: We don't have $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ but samples

- **Samples:** $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mu$ and $Y_1, \dots, Y_n \stackrel{\text{i.i.d.}}{\sim} \nu$
- **Empirical dist.:** $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ and $\nu_n := \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$

⇒ Can we approximate $W_p(\mu, \nu) \approx W_p(\mu_n, \nu_n)$?



Theorem (Dudley'69, Boissard-Le Gouic'14, Niles Weed-Bach'19,..)

For $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ and $d > 2p$: $\mathbb{E}[|W_p(\mu, \nu) - W_p(\mu_n, \nu_n)|] \asymp n^{-\frac{1}{d}}$

Core issue: For $\mu \ll \text{Leb}(\mathbb{R}^d)$: $W_p(\mu_n, \mu) \gtrsim n^{-\frac{1}{d}}$ a.s.

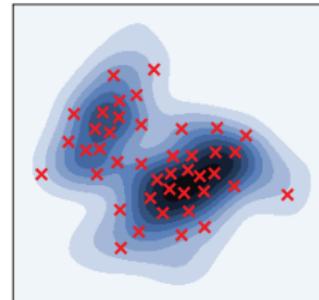
Curse of Dimensionality

Empirical Approximation in High Dimensions

In practice: We don't have $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ but samples

- **Samples:** $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mu$ and $Y_1, \dots, Y_n \stackrel{\text{i.i.d.}}{\sim} \nu$
- **Empirical dist.:** $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ and $\nu_n := \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$

⇒ Can we approximate $W_p(\mu, \nu) \approx W_p(\mu_n, \nu_n)$?



Theorem (Dudley'69, Boissard-Le Gouic'14, Niles Weed-Bach'19,..)

For $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ and $d > 2p$: $\mathbb{E}[|W_p(\mu, \nu) - W_p(\mu_n, \nu_n)|] \asymp n^{-\frac{1}{d}}$

Core issue: For $\mu \ll \text{Leb}(\mathbb{R}^d)$: $W_p(\mu_n, \mu) \gtrsim n^{-\frac{1}{d}}$ a.s.

✳ **Implication:** Too slow for $d \gg 1$ (real-world data)!

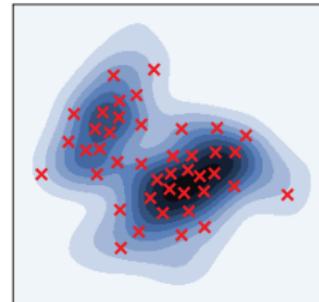
Curse of Dimensionality

Empirical Approximation in High Dimensions

In practice: We don't have $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ but samples

- **Samples:** $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mu$ and $Y_1, \dots, Y_n \stackrel{\text{i.i.d.}}{\sim} \nu$
- **Empirical dist.:** $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$ and $\nu_n := \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$

⇒ Can we approximate $W_p(\mu, \nu) \approx W_p(\mu_n, \nu_n)$?



Theorem (Dudley'69, Boissard-Le Gouic'14, Niles Weed-Bach'19,..)

For $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ and $d > 2p$: $\mathbb{E}[|W_p(\mu, \nu) - W_p(\mu_n, \nu_n)|] \asymp n^{-\frac{1}{d}}$

Core issue: For $\mu \ll \text{Leb}(\mathbb{R}^d)$: $W_p(\mu_n, \mu) \gtrsim n^{-\frac{1}{d}}$ a.s.

✳ **Implication:** Too slow for $d \gg 1$ (real-world data)!

✳ **Question:** How to preserve Wasserstein structure but alleviate CoD?

Curse of Dimensionality

Smooth 1-Wasserstein Distance

Definition (ZG-Greenewald'20)

For $\sigma \geq 0$, the smooth 1-Wasserstein distance between μ and ν is

$$W_1^\sigma(\mu, \nu) := W_1(\mu * \gamma_\sigma, \nu * \gamma_\sigma),$$

where $\gamma_\sigma := \mathcal{N}(0, \sigma^2 I_d)$ is a d -dimensional isotropic Gaussian.

Smooth 1-Wasserstein Distance

Definition (ZG-Greenewald'20)

For $\sigma \geq 0$, the smooth 1-Wasserstein distance between μ and ν is

$$W_1^\sigma(\mu, \nu) := W_1(\mu * \gamma_\sigma, \nu * \gamma_\sigma),$$

where $\gamma_\sigma := \mathcal{N}(0, \sigma^2 I_d)$ is a d -dimensional isotropic Gaussian.

Interpretation: $X \sim \mu$, $Y \sim \nu$ and $Z_1, Z_2 \sim \gamma_\sigma$

Smooth 1-Wasserstein Distance

Definition (ZG-Greenewald'20)

For $\sigma \geq 0$, the smooth 1-Wasserstein distance between μ and ν is

$$W_1^\sigma(\mu, \nu) := W_1(\mu * \gamma_\sigma, \nu * \gamma_\sigma),$$

where $\gamma_\sigma := \mathcal{N}(0, \sigma^2 I_d)$ is a d -dimensional isotropic Gaussian.

Interpretation: $X \sim \mu$, $Y \sim \nu$ and $Z_1, Z_2 \sim \gamma_\sigma$

$$X \perp Z_1 \implies X + Z_1 \sim \mu * \gamma_\sigma \quad \& \quad Y \perp Z_2 \implies Y + Z_2 \sim \nu * \gamma_\sigma$$

Smooth 1-Wasserstein Distance

Definition (ZG-Greenewald'20)

For $\sigma \geq 0$, the smooth 1-Wasserstein distance between μ and ν is

$$W_1^\sigma(\mu, \nu) := W_1(\mu * \gamma_\sigma, \nu * \gamma_\sigma),$$

where $\gamma_\sigma := \mathcal{N}(0, \sigma^2 I_d)$ is a d -dimensional isotropic Gaussian.

Interpretation: $X \sim \mu$, $Y \sim \nu$ and $Z_1, Z_2 \sim \gamma_\sigma$

$$X \perp Z_1 \implies X + Z_1 \sim \mu * \gamma_\sigma \quad \& \quad Y \perp Z_2 \implies Y + Z_2 \sim \nu * \gamma_\sigma$$

Structural properties: Wasserstein structure is largely preserved

Smooth 1-Wasserstein Distance

Definition (ZG-Greenewald'20)

For $\sigma \geq 0$, the smooth 1-Wasserstein distance between μ and ν is

$$W_1^\sigma(\mu, \nu) := W_1(\mu * \gamma_\sigma, \nu * \gamma_\sigma),$$

where $\gamma_\sigma := \mathcal{N}(0, \sigma^2 I_d)$ is a d -dimensional isotropic Gaussian.

Interpretation: $X \sim \mu$, $Y \sim \nu$ and $Z_1, Z_2 \sim \gamma_\sigma$

$$X \perp Z_1 \implies X + Z_1 \sim \mu * \gamma_\sigma \quad \& \quad Y \perp Z_2 \implies Y + Z_2 \sim \nu * \gamma_\sigma$$

Structural properties: Wasserstein structure is largely preserved

① **Metric:** $(\mathcal{P}_1(\mathbb{R}^d), W_1^\sigma)$ metric space & topologically equivalent to W_1

Smooth 1-Wasserstein Distance

Definition (ZG-Greenewald'20)

For $\sigma \geq 0$, the smooth 1-Wasserstein distance between μ and ν is

$$W_1^\sigma(\mu, \nu) := W_1(\mu * \gamma_\sigma, \nu * \gamma_\sigma),$$

where $\gamma_\sigma := \mathcal{N}(0, \sigma^2 I_d)$ is a d -dimensional isotropic Gaussian.

Interpretation: $X \sim \mu$, $Y \sim \nu$ and $Z_1, Z_2 \sim \gamma_\sigma$

$$X \perp Z_1 \implies X + Z_1 \sim \mu * \gamma_\sigma \quad \& \quad Y \perp Z_2 \implies Y + Z_2 \sim \nu * \gamma_\sigma$$

Structural properties: Wasserstein structure is largely preserved

- ① **Metric:** $(\mathcal{P}_1(\mathbb{R}^d), W_1^\sigma)$ metric space & topologically equivalent to W_1
- ② **KR duality:** $W_1^\sigma(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \int f \, d(\mu - \nu) * \gamma_\sigma$

Smooth 1-Wasserstein Distance

Definition (ZG-Greenewald'20)

For $\sigma \geq 0$, the smooth 1-Wasserstein distance between μ and ν is

$$W_1^\sigma(\mu, \nu) := W_1(\mu * \gamma_\sigma, \nu * \gamma_\sigma),$$

where $\gamma_\sigma := \mathcal{N}(0, \sigma^2 I_d)$ is a d -dimensional isotropic Gaussian.

Interpretation: $X \sim \mu$, $Y \sim \nu$ and $Z_1, Z_2 \sim \gamma_\sigma$

$$X \perp Z_1 \implies X + Z_1 \sim \mu * \gamma_\sigma \quad \& \quad Y \perp Z_2 \implies Y + Z_2 \sim \nu * \gamma_\sigma$$

Structural properties: Wasserstein structure is largely preserved

- ① **Metric:** $(\mathcal{P}_1(\mathbb{R}^d), W_1^\sigma)$ metric space & topologically equivalent to W_1
- ② **KR duality:** $W_1^\sigma(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \int f \, d(\mu - \nu) * \gamma_\sigma$
- ③ **Stability:** $|W_1^\sigma(\mu, \nu) - W_1^\tau(\mu, \nu)| \leq 2\sqrt{d|\sigma^2 - \tau^2|}$

Smooth 1-Wasserstein Distance

Definition (ZG-Greenewald'20)

For $\sigma \geq 0$, the smooth 1-Wasserstein distance between μ and ν is

$$W_1^\sigma(\mu, \nu) := W_1(\mu * \gamma_\sigma, \nu * \gamma_\sigma),$$

where $\gamma_\sigma := \mathcal{N}(0, \sigma^2 I_d)$ is a d -dimensional isotropic Gaussian.

Interpretation: $X \sim \mu$, $Y \sim \nu$ and $Z_1, Z_2 \sim \gamma_\sigma$

$$X \perp Z_1 \implies X + Z_1 \sim \mu * \gamma_\sigma \quad \& \quad Y \perp Z_2 \implies Y + Z_2 \sim \nu * \gamma_\sigma$$

Structural properties: Wasserstein structure is largely preserved

- ① **Metric:** $(\mathcal{P}_1(\mathbb{R}^d), W_1^\sigma)$ metric space & topologically equivalent to W_1
- ② **KR duality:** $W_1^\sigma(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \int f \, d(\mu - \nu) * \gamma_\sigma$
- ③ **Stability:** $|W_1^\sigma(\mu, \nu) - W_1^\tau(\mu, \nu)| \leq 2\sqrt{d|\sigma^2 - \tau^2|}$
- ④ **Regularity:** Cont. & monotone func. of σ + convergence of plans

Smooth 1-Wasserstein: Statistical Aspects

High level: Parametric convergence & high-dimensional limit distribution

Smooth 1-Wasserstein: Statistical Aspects

High level: Parametric convergence & high-dimensional limit distribution

Theorem (ZG-Greenwald-Polyanskiy-Weed'20, Sadhu-ZG-Kato'21)

For any $d \geq 1$, $\sigma > 0$ and sub-Gaussian μ : $\mathbb{E}[W_1^\sigma(\mu_n, \mu)] \lesssim \sigma^{-\frac{d}{2}+1} n^{-\frac{1}{2}}$

Smooth 1-Wasserstein: Statistical Aspects

High level: Parametric convergence & high-dimensional limit distribution

Theorem (ZG-Greenwald-Polyanskiy-Weed'20, Sadhu-ZG-Kato'21)

For any $d \geq 1$, $\sigma > 0$ and sub-Gaussian μ : $\mathbb{E}[W_1^\sigma(\mu_n, \mu)] \lesssim \sigma^{-\frac{d}{2}+1} n^{-\frac{1}{2}}$

Theorem (ZG-Kato'20, Sadhu-ZG-Kato'21)

Under same assumptions:

$$\sqrt{n} W_1^\sigma(\mu_n, \mu) \xrightarrow{d} \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} G_\mu^\sigma(f),$$

where G_μ^σ is a tight centered Gaussian process in $\ell^\infty(\text{Lip}_1(\mathbb{R}^d))$ with

$$\mathbb{E}[G_\mu^\sigma(f) G_\mu^\sigma(g)] = \text{Cov}_\mu(f * \varphi_\sigma, g * \varphi_\sigma), \quad \forall f, g \in \text{Lip}_1(\mathbb{R}^d)$$

Smooth 1-Wasserstein: Statistical Aspects

High level: Parametric convergence & high-dimensional limit distribution

Theorem (ZG-Greenwald-Polyanskiy-Weed'20, Sadhu-ZG-Kato'21)

For any $d \geq 1$, $\sigma > 0$ and sub-Gaussian μ : $\mathbb{E}[W_1^\sigma(\mu_n, \mu)] \lesssim \sigma^{-\frac{d}{2}+1} n^{-\frac{1}{2}}$

Theorem (ZG-Kato'20, Sadhu-ZG-Kato'21)

Under same assumptions:

$$\sqrt{n}W_1^\sigma(\mu_n, \mu) \xrightarrow{d} \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} G_\mu^\sigma(f),$$

where G_μ^σ is a tight centered Gaussian process in $\ell^\infty(\text{Lip}_1(\mathbb{R}^d))$ with
 $\mathbb{E}[G_\mu^\sigma(f)G_\mu^\sigma(g)] = \text{Cov}_\mu(f * \varphi_\sigma, g * \varphi_\sigma), \quad \forall f, g \in \text{Lip}_1(\mathbb{R}^d)$

Comments:

- W_1 limit dist. known only for $d = 1$ ($W_1(\mu_n, \mu) = \|F_n - F\|_{L^1(\mathbb{R})}$)

Smooth 1-Wasserstein: Statistical Aspects

High level: Parametric convergence & high-dimensional limit distribution

Theorem (ZG-Greenwald-Polyanskiy-Weed'20, Sadhu-ZG-Kato'21)

For any $d \geq 1$, $\sigma > 0$ and sub-Gaussian μ : $\mathbb{E}[W_1^\sigma(\mu_n, \mu)] \lesssim \sigma^{-\frac{d}{2}+1} n^{-\frac{1}{2}}$

Theorem (ZG-Kato'20, Sadhu-ZG-Kato'21)

Under same assumptions:

$$\sqrt{n}W_1^\sigma(\mu_n, \mu) \xrightarrow{d} \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} G_\mu^\sigma(f),$$

where G_μ^σ is a tight centered Gaussian process in $\ell^\infty(\text{Lip}_1(\mathbb{R}^d))$ with
 $\mathbb{E}[G_\mu^\sigma(f)G_\mu^\sigma(g)] = \text{Cov}_\mu(f * \varphi_\sigma, g * \varphi_\sigma), \quad \forall f, g \in \text{Lip}_1(\mathbb{R}^d)$

Comments:

- W_1 limit dist. known only for $d = 1$ ($W_1(\mu_n, \mu) = \|F_n - F\|_{L^1(\mathbb{R})}$)
- **Extensions:** One- and two-sample cases under null and alternative

Limit Distribution Proof Outline

- ① KR duality: Let $\mathcal{F}_\sigma := \{f * \varphi_\sigma : f \in \text{Lip}_1(\mathbb{R}^d)\}$ and write

Limit Distribution Proof Outline

- ① KR duality: Let $\mathcal{F}_\sigma := \{f * \varphi_\sigma : f \in \text{Lip}_1(\mathbb{R}^d)\}$ and write

$$W_1^\sigma(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \int f * \varphi_\sigma \, d(\mu - \nu) = \underbrace{\sup_{g \in \mathcal{F}_\sigma} \int g \, d(\mu - \nu)}_{\text{IPM wrt } \mathcal{F}_\sigma}$$

Limit Distribution Proof Outline

① **KR duality:** Let $\mathcal{F}_\sigma := \{f * \varphi_\sigma : f \in \text{Lip}_1(\mathbb{R}^d)\}$ and write

$$W_1^\sigma(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \int f * \varphi_\sigma \, d(\mu - \nu) = \underbrace{\sup_{g \in \mathcal{F}_\sigma} \int g \, d(\mu - \nu)}_{\text{IPM wrt } \mathcal{F}_\sigma}$$

\implies **Emp.** W_1^σ :

$$\sqrt{n} W_1^\sigma(\mu_n, \mu) = \sqrt{n} \sup_{g \in \mathcal{F}_\sigma} \underbrace{\frac{1}{n} \sum_{i=1}^n g(X_i) - \mathbb{E}[g(X)]}_{=: (\mu_n - \mu)g}$$

Limit Distribution Proof Outline

① **KR duality:** Let $\mathcal{F}_\sigma := \{f * \varphi_\sigma : f \in \text{Lip}_1(\mathbb{R}^d)\}$ and write

$$W_1^\sigma(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \int f * \varphi_\sigma \, d(\mu - \nu) = \underbrace{\sup_{g \in \mathcal{F}_\sigma} \int g \, d(\mu - \nu)}_{\text{IPM wrt } \mathcal{F}_\sigma}$$

⇒ **Emp. W_1^σ :**

$$\sqrt{n} W_1^\sigma(\mu_n, \mu) = \sqrt{n} \sup_{g \in \mathcal{F}_\sigma} \underbrace{\frac{1}{n} \sum_{i=1}^n g(X_i) - \mathbb{E}[g(X)]}_{=: (\mu_n - \mu)g}$$

* **Objective:** $\mathbb{G}_n = (\sqrt{n}(\mu_n - \mu)g)_{g \in \mathcal{F}_\sigma} \xrightarrow{w} \text{GP}$

Limit Distribution Proof Outline

① **KR duality:** Let $\mathcal{F}_\sigma := \{f * \varphi_\sigma : f \in \text{Lip}_1(\mathbb{R}^d)\}$ and write

$$W_1^\sigma(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \int f * \varphi_\sigma \, d(\mu - \nu) = \underbrace{\sup_{g \in \mathcal{F}_\sigma} \int g \, d(\mu - \nu)}_{\text{IPM wrt } \mathcal{F}_\sigma}$$

⇒ **Emp. W_1^σ :**

$$\sqrt{n} W_1^\sigma(\mu_n, \mu) = \sqrt{n} \sup_{g \in \mathcal{F}_\sigma} \underbrace{\frac{1}{n} \sum_{i=1}^n g(X_i) - \mathbb{E}[g(X)]}_{=: (\mu_n - \mu)g}$$

* **Objective:** $\mathbb{G}_n = (\sqrt{n}(\mu_n - \mu)g)_{g \in \mathcal{F}_\sigma} \xrightarrow{w} \text{GP} \iff \mathcal{F}_\sigma \text{ is } \mu - \text{Donsker}$

Limit Distribution Proof Outline

① **KR duality:** Let $\mathcal{F}_\sigma := \{f * \varphi_\sigma : f \in \text{Lip}_1(\mathbb{R}^d)\}$ and write

$$W_1^\sigma(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \int f * \varphi_\sigma \, d(\mu - \nu) = \underbrace{\sup_{g \in \mathcal{F}_\sigma} \int g \, d(\mu - \nu)}_{\text{IPM wrt } \mathcal{F}_\sigma}$$

\implies **Emp. W_1^σ :**

$$\sqrt{n} W_1^\sigma(\mu_n, \mu) = \sqrt{n} \sup_{g \in \mathcal{F}_\sigma} \underbrace{\frac{1}{n} \sum_{i=1}^n g(X_i) - \mathbb{E}[g(X)]}_{=: (\mu_n - \mu)g}$$

② **Objective:** $\mathbb{G}_n = (\sqrt{n}(\mu_n - \mu)g)_{g \in \mathcal{F}_\sigma} \xrightarrow{w} \text{GP} \iff \mathcal{F}_\sigma \text{ is } \mu - \text{Donsker}$

③ **Uniform derivative bound:** $\|D^k g\|_\infty \leq \sigma^{-|k|+1} \sqrt{(|k|-1)!}, \quad \forall g \in \mathcal{F}_\sigma$

Limit Distribution Proof Outline

① **KR duality:** Let $\mathcal{F}_\sigma := \{f * \varphi_\sigma : f \in \text{Lip}_1(\mathbb{R}^d)\}$ and write

$$W_1^\sigma(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \int f * \varphi_\sigma \, d(\mu - \nu) = \underbrace{\sup_{g \in \mathcal{F}_\sigma} \int g \, d(\mu - \nu)}_{\text{IPM wrt } \mathcal{F}_\sigma}$$

\implies **Emp. W_1^σ :**

$$\sqrt{n} W_1^\sigma(\mu_n, \mu) = \sqrt{n} \sup_{g \in \mathcal{F}_\sigma} \underbrace{\frac{1}{n} \sum_{i=1}^n g(X_i) - \mathbb{E}[g(X)]}_{=: (\mu_n - \mu)g}$$

② **Objective:** $\mathbb{G}_n = (\sqrt{n}(\mu_n - \mu)g)_{g \in \mathcal{F}_\sigma} \xrightarrow{w} \text{GP} \iff \mathcal{F}_\sigma \text{ is } \mu - \text{Donsker}$

③ **Uniform derivative bound:** $\|D^k g\|_\infty \leq \sigma^{-|k|+1} \sqrt{(|k|-1)!}, \quad \forall g \in \mathcal{F}_\sigma$

$\implies \mathcal{F}_\sigma \text{ is } \mu\text{-Donsker:}$

$\mathbb{G}_n \xrightarrow{w} G_\mu$ (Gaussian process)

Limit Distribution Proof Outline

- ① **KR duality:** Let $\mathcal{F}_\sigma := \{f * \varphi_\sigma : f \in \text{Lip}_1(\mathbb{R}^d)\}$ and write

$$W_1^\sigma(\mu, \nu) = \sup_{f \in \text{Lip}_1(\mathbb{R}^d)} \int f * \varphi_\sigma \, d(\mu - \nu) = \underbrace{\sup_{g \in \mathcal{F}_\sigma} \int g \, d(\mu - \nu)}_{\text{IPM wrt } \mathcal{F}_\sigma}$$

\implies **Emp.** W_1^σ :

$$\sqrt{n} W_1^\sigma(\mu_n, \mu) = \sqrt{n} \sup_{g \in \mathcal{F}_\sigma} \underbrace{\frac{1}{n} \sum_{i=1}^n g(X_i) - \mathbb{E}[g(X)]}_{=: (\mu_n - \mu)g}$$

- ② **Objective:** $\mathbb{G}_n = (\sqrt{n}(\mu_n - \mu)g)_{g \in \mathcal{F}_\sigma} \xrightarrow{w} \text{GP} \iff \mathcal{F}_\sigma \text{ is } \mu - \text{Donsker}$

- ③ **Uniform derivative bound:** $\|D^k g\|_\infty \leq \sigma^{-|k|+1} \sqrt{(|k|-1)!}, \quad \forall g \in \mathcal{F}_\sigma$

$\implies \mathcal{F}_\sigma$ is μ -Donsker:

$\mathbb{G}_n \xrightarrow{w} G_\mu$ (Gaussian process)

- ④ **Continuous mapping theorem:**

$\sqrt{n} W_1^\sigma(\mu_n, \mu) \xrightarrow{d} \sup_{g \in \mathcal{F}_\sigma} G_\mu(g)$

Smooth p -Wasserstein Distances

Definition (Smooth p -Wasserstein distance)

For $\sigma \geq 0$ and $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, define $W_p^\sigma(\mu, \nu) := W_p(\mu * \gamma_\sigma, \nu * \gamma_\sigma)$.

Smooth p -Wasserstein Distances

Definition (Smooth p -Wasserstein distance)

For $\sigma \geq 0$ and $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, define $W_p^\sigma(\mu, \nu) := W_p(\mu * \gamma_\sigma, \nu * \gamma_\sigma)$.

Structure: Everything carries over (metric/topology/stability/regularity)

Smooth p -Wasserstein Distances

Definition (Smooth p -Wasserstein distance)

For $\sigma \geq 0$ and $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, define $W_p^\sigma(\mu, \nu) := W_p(\mu * \gamma_\sigma, \nu * \gamma_\sigma)$.

Structure: Everything carries over (metric/topology/stability/regularity)

Statistical aspects: Fast empirical convergence & limit distributions?

Smooth p -Wasserstein Distances

Definition (Smooth p -Wasserstein distance)

For $\sigma \geq 0$ and $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, define $W_p^\sigma(\mu, \nu) := W_p(\mu * \gamma_\sigma, \nu * \gamma_\sigma)$.

Structure: Everything carries over (metric/topology/stability/regularity)

Statistical aspects: Fast empirical convergence & limit distributions?

- **Recall:** $W_1^\sigma(\mu_n, \mu) = \sup_{g \in \mathcal{F}_\sigma} (\mu_n - \mu)(g)$ IPM structure key for analysis

Smooth p -Wasserstein Distances

Definition (Smooth p -Wasserstein distance)

For $\sigma \geq 0$ and $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, define $W_p^\sigma(\mu, \nu) := W_p(\mu * \gamma_\sigma, \nu * \gamma_\sigma)$.

Structure: Everything carries over (metric/topology/stability/regularity)

Statistical aspects: Fast empirical convergence & limit distributions?

- **Recall:** $W_1^\sigma(\mu_n, \mu) = \sup_{g \in \mathcal{F}_\sigma} (\mu_n - \mu)(g)$ IPM structure key for analysis
- ✳ **Issue:** W_p is not an IPM for $p > 1$

Smooth p -Wasserstein Distances

Definition (Smooth p -Wasserstein distance)

For $\sigma \geq 0$ and $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$, define $W_p^\sigma(\mu, \nu) := W_p(\mu * \gamma_\sigma, \nu * \gamma_\sigma)$.

Structure: Everything carries over (metric/topology/stability/regularity)

Statistical aspects: Fast empirical convergence & limit distributions?

- **Recall:** $W_1^\sigma(\mu_n, \mu) = \sup_{g \in \mathcal{F}_\sigma} (\mu_n - \mu)(g)$ IPM structure key for analysis

- ✳ **Issue:** W_p is not an IPM for $p > 1$

- ✳ **Strategy:** Relate W_p to an IPM & leverage proof techniques

Which IPM? Dual Sobolev Norm

Setup: Sobolev space and dual space

Which IPM? Dual Sobolev Norm

Setup: Sobolev space and dual space

- **Sobolev seminorm:** $\|f\|_{\dot{H}^{1,p}(\rho)} := \|\nabla f\|_{L^p(\rho; \mathbb{R}^d)} = (\int_{\mathbb{R}^d} \|\nabla f\|^p d\rho)^{1/p}$

Which IPM? Dual Sobolev Norm

Setup: Sobolev space and dual space

- **Sobolev seminorm:** $\|f\|_{\dot{H}^{1,p}(\rho)} := \|\nabla f\|_{L^p(\rho; \mathbb{R}^d)} = (\int_{\mathbb{R}^d} \|\nabla f\|^p d\rho)^{1/p}$
- **Homogen. Sobolev space:** $\dot{H}^{1,p}(\rho) = \text{completion of } C_0^\infty \text{ wrt } \|\cdot\|_{\dot{H}^{1,p}(\rho)}$

Which IPM? Dual Sobolev Norm

Setup: Sobolev space and dual space

- **Sobolev seminorm:** $\|f\|_{\dot{H}^{1,p}(\rho)} := \|\nabla f\|_{L^p(\rho; \mathbb{R}^d)} = (\int_{\mathbb{R}^d} \|\nabla f\|^p d\rho)^{1/p}$
- **Homogen. Sobolev space:** $\dot{H}^{1,p}(\rho) = \text{completion of } C_0^\infty \text{ wrt } \|\cdot\|_{\dot{H}^{1,p}(\rho)}$
- **Dual space:** $\dot{H}^{-1,q}(\rho) = \text{topological dual of } \dot{H}^{1,p}(\rho) \left(\frac{1}{p} + \frac{1}{q} = 1 \right) \text{ and}$

$$\|\ell\|_{\dot{H}^{-1,q}(\rho)} := \sup_{\substack{f \in C_0^\infty: \\ \|f\|_{\dot{H}^{1,p}(\rho)} \leq 1}} \ell(f)$$

Which IPM? Dual Sobolev Norm

Setup: Sobolev space and dual space

- **Sobolev seminorm:** $\|f\|_{\dot{H}^{1,p}(\rho)} := \|\nabla f\|_{L^p(\rho; \mathbb{R}^d)} = (\int_{\mathbb{R}^d} \|\nabla f\|^p d\rho)^{1/p}$
- **Homogen. Sobolev space:** $\dot{H}^{1,p}(\rho) = \text{completion of } C_0^\infty \text{ wrt } \|\cdot\|_{\dot{H}^{1,p}(\rho)}$
- **Dual space:** $\dot{H}^{-1,q}(\rho) = \text{topological dual of } \dot{H}^{1,p}(\rho) \left(\frac{1}{p} + \frac{1}{q} = 1 \right) \text{ and}$

$$\|\ell\|_{\dot{H}^{-1,q}(\rho)} := \sup_{\substack{f \in C_0^\infty: \\ \|f\|_{\dot{H}^{1,p}(\rho)} \leq 1}} \ell(f)$$

Identify: Finite signed Borel measure ℓ on \mathbb{R}^d w/ lin. func. $\ell(f) := \int f d\ell$

Which IPM? Dual Sobolev Norm

Setup: Sobolev space and dual space

- **Sobolev seminorm:** $\|f\|_{\dot{H}^{1,p}(\rho)} := \|\nabla f\|_{L^p(\rho; \mathbb{R}^d)} = (\int_{\mathbb{R}^d} \|\nabla f\|^p d\rho)^{1/p}$
- **Homogen. Sobolev space:** $\dot{H}^{1,p}(\rho) = \text{completion of } C_0^\infty \text{ wrt } \|\cdot\|_{\dot{H}^{1,p}(\rho)}$
- **Dual space:** $\dot{H}^{-1,q}(\rho) = \text{topological dual of } \dot{H}^{1,p}(\rho) \left(\frac{1}{p} + \frac{1}{q} = 1 \right) \text{ and}$

$$\|\ell\|_{\dot{H}^{-1,q}(\rho)} := \sup_{\substack{f \in C_0^\infty: \\ \|f\|_{\dot{H}^{1,p}(\rho)} \leq 1}} \ell(f)$$

Identify: Finite signed Borel measure ℓ on \mathbb{R}^d w/ lin. func. $\ell(f) := \int f d\ell$

Definition (Sobolev IPM = dual Sobolev norm)

For $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$: $\|\mu - \nu\|_{\dot{H}^{-1,p}(\rho)} = \sup_{\substack{f \in C_0^\infty: \\ \|f\|_{\dot{H}^{1,q}(\rho)} \leq 1}} \int f d(\mu - \nu).$

Which IPM? Dual Sobolev Norm

Setup: Sobolev space and dual space

- **Sobolev seminorm:** $\|f\|_{\dot{H}^{1,p}(\rho)} := \|\nabla f\|_{L^p(\rho; \mathbb{R}^d)} = (\int_{\mathbb{R}^d} \|\nabla f\|^p d\rho)^{1/p}$
- **Homogen. Sobolev space:** $\dot{H}^{1,p}(\rho) = \text{completion of } C_0^\infty \text{ wrt } \|\cdot\|_{\dot{H}^{1,p}(\rho)}$
- **Dual space:** $\dot{H}^{-1,q}(\rho) = \text{topological dual of } \dot{H}^{1,p}(\rho) \left(\frac{1}{p} + \frac{1}{q} = 1 \right) \text{ and}$

$$\|\ell\|_{\dot{H}^{-1,q}(\rho)} := \sup_{\substack{f \in C_0^\infty: \\ \|f\|_{\dot{H}^{1,p}(\rho)} \leq 1}} \ell(f)$$

Identify: Finite signed Borel measure ℓ on \mathbb{R}^d w/ lin. func. $\ell(f) := \int f d\ell$

Definition (Sobolev IPM = dual Sobolev norm)

For $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$: $\|\mu - \nu\|_{\dot{H}^{-1,p}(\rho)} = \sup_{\substack{f \in C_0^\infty: \\ \|f\|_{\dot{H}^{1,q}(\rho)} \leq 1}} \int f d(\mu - \nu).$

✳ **Question:** How do Wasserstein dist. & dual Sobolev norm relate?

Smooth p -Wasserstein vs Dual Sobolev Norm

Theorem (Dolbeault-Nazaret-Savaré'09)

Let $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ satisfy $\mu, \nu \ll \rho$ with $\frac{d\mu}{d\rho} \geq c > 0$. Then

$$W_p(\mu, \nu) \leq pc^{-1/q} \|\mu - \nu\|_{\dot{H}^{-1,p}(\rho)}.$$

Smooth p -Wasserstein vs Dual Sobolev Norm

Theorem (Dolbeault-Nazaret-Savaré'09)

Let $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ satisfy $\mu, \nu \ll \rho$ with $\frac{d\mu}{d\rho} \geq c > 0$. Then

$$W_p(\mu, \nu) \leq pc^{-1/q} \|\mu - \nu\|_{\dot{H}^{-1,p}(\rho)}.$$

Comments:

Smooth p -Wasserstein vs Dual Sobolev Norm

Theorem (Dolbeault-Nazaret-Savaré'09)

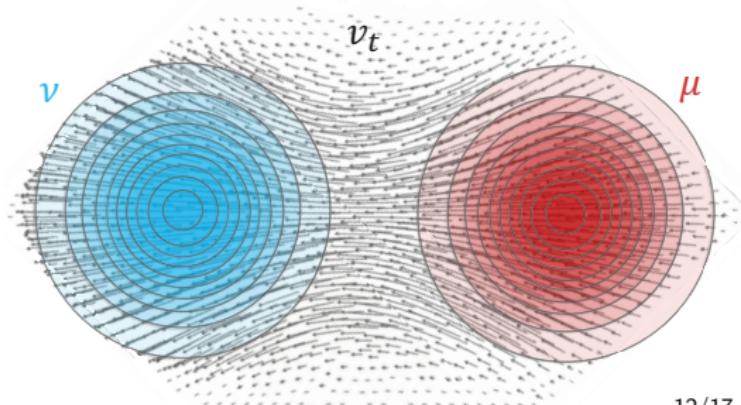
Let $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ satisfy $\mu, \nu \ll \rho$ with $\frac{d\mu}{d\rho} \geq c > 0$. Then

$$W_p(\mu, \nu) \leq pc^{-1/q} \|\mu - \nu\|_{\dot{H}^{-1,p}(\rho)}.$$

Comments:

- 1 Proof via Benamou-Brenier dynamical formulation

$$W_p(\mu, \nu) = \inf_{\mu_t, v_t} \left\{ \int_0^1 \|v_t\|_{L^p(\mu_t)} dt : \mu_0 = \mu, \mu_1 = \nu, \partial_t \mu_t + \nabla \cdot (\mu_t v_t) = 0 \right\}$$



Smooth p -Wasserstein vs Dual Sobolev Norm

Theorem (Dolbeault-Nazaret-Savaré'09)

Let $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ satisfy $\mu, \nu \ll \rho$ with $\frac{d\mu}{d\rho} \geq c > 0$. Then

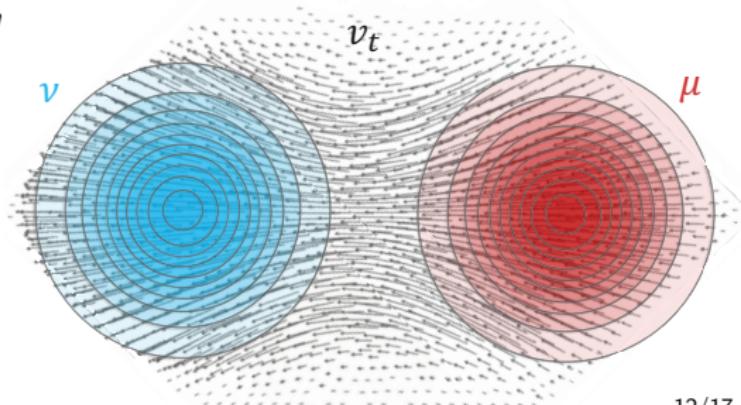
$$W_p(\mu, \nu) \leq pc^{-1/q} \|\mu - \nu\|_{\dot{H}^{-1,p}(\rho)}.$$

Comments:

- 1 Proof via Benamou-Brenier dynamical formulation

$$W_p(\mu, \nu) = \inf_{\mu_t, v_t} \left\{ \int_0^1 \|v_t\|_{L^p(\mu_t)} dt : \mu_0 = \mu, \mu_1 = \nu, \partial_t \mu_t + \nabla \cdot (\mu_t v_t) = 0 \right\}$$

- Take $\mu_t = (1-t)\mu + t\nu$ & $v_t = v$



Smooth p -Wasserstein vs Dual Sobolev Norm

Theorem (Dolbeault-Nazaret-Savaré'09)

Let $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ satisfy $\mu, \nu \ll \rho$ with $\frac{d\mu}{d\rho} \geq c > 0$. Then

$$W_p(\mu, \nu) \leq pc^{-1/q} \|\mu - \nu\|_{\dot{H}^{-1,p}(\rho)}.$$

Comments:

- 1 Proof via Benamou-Brenier dynamical formulation

$$W_p(\mu, \nu) = \inf_{\mu_t, v_t} \left\{ \int_0^1 \|v_t\|_{L^p(\mu_t)} dt : \mu_0 = \mu, \mu_1 = \nu, \partial_t \mu_t + \nabla \cdot (\mu_t v_t) = 0 \right\}$$

- Take $\mu_t = (1-t)\mu + t\nu$ & $v_t = v$

- 2 Lower bound automatically holds after Gaussian smoothing w/ $\rho = \gamma_\sigma$

Smooth p -Wasserstein vs Dual Sobolev Norm

Theorem (Dolbeault-Nazaret-Savaré'09)

Let $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$ satisfy $\mu, \nu \ll \rho$ with $\frac{d\mu}{d\rho} \geq c > 0$. Then

$$W_p(\mu, \nu) \leq pc^{-1/q} \|\mu - \nu\|_{\dot{H}^{-1,p}(\rho)}.$$

Comments:

- 1 Proof via Benamou-Brenier dynamical formulation

$$W_p(\mu, \nu) = \inf_{\mu_t, v_t} \left\{ \int_0^1 \|v_t\|_{L^p(\mu_t)} dt : \mu_0 = \mu, \mu_1 = \nu, \partial_t \mu_t + \nabla \cdot (\mu_t v_t) = 0 \right\}$$

► Take $\mu_t = (1-t)\mu + t\nu$ & $v_t = v$

- 2 Lower bound automatically holds after Gaussian smoothing w/ $\rho = \gamma_\sigma$

Theorem (Nietert-ZG-Kato'21)

For $X \sim \mu \in \mathcal{P}_{p\vee 2}(\mathbb{R}^d)$ with mean 0 and $\nu \in \mathcal{P}(\mathbb{R}^d)$,

$$W_p^\sigma(\mu, \nu) \leq p e^{\mathbb{E}[\|X\|^2]/(2q\sigma^2)} \|(\mu - \nu) * \gamma_\sigma\|_{\dot{H}^{-1,p}(\gamma_\sigma)}.$$

Smooth p -Wasserstein: Fast Empirical Convergence

Theorem (Nietert-ZG-Kato'21)

For μ β -sub-Gaussian with $\beta < \sigma / \sqrt{2(p-1)}$: $\mathbb{E}[W_p^\sigma(\mu_n, \mu)] \lesssim n^{-\frac{1}{2}}$.

Smooth p -Wasserstein: Fast Empirical Convergence

Theorem (Nietert-ZG-Kato'21)

For μ β -sub-Gaussian with $\beta < \sigma/\sqrt{2(p-1)}$: $\mathbb{E}[W_p^\sigma(\mu_n, \mu)] \lesssim n^{-\frac{1}{2}}$.

Pf Idea: Hinges on comparison $W_p^\sigma(\mu, \nu) \lesssim_{p,\sigma} \|(\mu - \nu) * \gamma_\sigma\|_{\dot{H}^{-1,p}(\gamma_\sigma)}$

Smooth p -Wasserstein: Fast Empirical Convergence

Theorem (Nietert-ZG-Kato'21)

For μ β -sub-Gaussian with $\beta < \sigma/\sqrt{2(p-1)}$: $\mathbb{E}[W_p^\sigma(\mu_n, \mu)] \lesssim n^{-\frac{1}{2}}$.

Pf Idea: Hinges on comparison $W_p^\sigma(\mu, \nu) \lesssim_{p,\sigma} \|(\mu - \nu) * \gamma_\sigma\|_{\dot{H}^{-1,p}(\gamma_\sigma)}$

- ① Bound derivatives of smooth Sobolev class $\{f \in C_0^\infty : \|f\|_{\dot{H}^{1,q}} \leq 1\} * \varphi_\sigma$

Smooth p -Wasserstein: Fast Empirical Convergence

Theorem (Nietert-ZG-Kato'21)

For μ β -sub-Gaussian with $\beta < \sigma/\sqrt{2(p-1)}$: $\mathbb{E}[W_p^\sigma(\mu_n, \mu)] \lesssim n^{-\frac{1}{2}}$.

Pf Idea: Hinges on comparison $W_p^\sigma(\mu, \nu) \lesssim_{p,\sigma} \|(\mu - \nu) * \gamma_\sigma\|_{\dot{H}^{-1,p}(\gamma_\sigma)}$

- ① Bound derivatives of smooth Sobolev class $\{f \in C_0^\infty : \|f\|_{\dot{H}^{1,q}} \leq 1\} * \varphi_\sigma$
→ Class is μ -Donsker (under sub-Gaussian cond.)

Smooth p -Wasserstein: Fast Empirical Convergence

Theorem (Nietert-ZG-Kato'21)

For μ β -sub-Gaussian with $\beta < \sigma/\sqrt{2(p-1)}$: $\mathbb{E}[W_p^\sigma(\mu_n, \mu)] \lesssim n^{-\frac{1}{2}}$.

Pf Idea: Hinges on comparison $W_p^\sigma(\mu, \nu) \lesssim_{p,\sigma} \|(\mu - \nu) * \gamma_\sigma\|_{\dot{H}^{-1,p}(\gamma_\sigma)}$

① Bound derivatives of smooth Sobolev class $\{f \in C_0^\infty : \|f\|_{\dot{H}^{1,q}} \leq 1\} * \varphi_\sigma$

→ Class is μ -Donsker (under sub-Gaussian cond.)

② Gives $\sqrt{n}\|(\mu_n - \mu) * \gamma_\sigma\|_{\dot{H}^{-1,p}(\gamma_\sigma)}$ limit distribution & sup-norm bound

$$\mathbb{E}\left[\|(\mu_n - \mu) * \gamma_\sigma\|_{\dot{H}^{-1,p}(\gamma_\sigma)}\right] \lesssim n^{-\frac{1}{2}}$$

Smooth p -Wasserstein: Fast Empirical Convergence

Theorem (Nietert-ZG-Kato'21)

For μ β -sub-Gaussian with $\beta < \sigma/\sqrt{2(p-1)}$: $\mathbb{E}[W_p^\sigma(\mu_n, \mu)] \lesssim n^{-\frac{1}{2}}$.

Pf Idea: Hinges on comparison $W_p^\sigma(\mu, \nu) \lesssim_{p,\sigma} \|(\mu - \nu) * \gamma_\sigma\|_{\dot{H}^{-1,p}(\gamma_\sigma)}$

- ➊ Bound derivatives of smooth Sobolev class $\{f \in C_0^\infty : \|f\|_{\dot{H}^{1,q}} \leq 1\} * \varphi_\sigma$
➡ Class is μ -Donsker (under sub-Gaussian cond.)
- ➋ Gives $\sqrt{n}\|(\mu_n - \mu) * \gamma_\sigma\|_{\dot{H}^{-1,p}(\gamma_\sigma)}$ limit distribution & sup-norm bound
$$\mathbb{E}\left[\|(\mu_n - \mu) * \gamma_\sigma\|_{\dot{H}^{-1,p}(\gamma_\sigma)}\right] \lesssim n^{-\frac{1}{2}}$$
- ➌ Apply comparison to obtain parametric empirical convergence for W_p^σ

Smooth p -Wasserstein: Limit Distribution

Approach: Extended functional delta method in normed spaces

Smooth p -Wasserstein: Limit Distribution

Approach: Extended functional delta method in normed spaces

Theorem (Shapiro'90, Römisch'04)

Let N be a normed vector space (NVS) and

- ① $(X_n)_{n \in \mathbb{N}} \subset D \subset N$ s.t. $\sqrt{n}(X_n - u) \xrightarrow{d} X$
- ② $\Phi : D \rightarrow \mathbb{R}$ is a functional
- ③ Φ is Hadamard directionally differentiable at u w/ derivative Φ'_u .

Then $\sqrt{n}(\Phi(X_n) - \Phi(u)) \xrightarrow{d} \Phi'_u(X)$.

Smooth p -Wasserstein: Limit Distribution

Approach: Extended functional delta method in normed spaces

Theorem (Shapiro'90, Römisch'04)

Let N be a normed vector space (NVS) and

- ① $(X_n)_{n \in \mathbb{N}} \subset D \subset N$ s.t. $\sqrt{n}(X_n - u) \xrightarrow{d} X$
- ② $\Phi : D \rightarrow \mathbb{R}$ is a functional
- ③ Φ is Hadamard directionally differentiable at u w/ derivative Φ'_u .

Then $\sqrt{n}(\Phi(X_n) - \Phi(u)) \xrightarrow{d} \Phi'_u(X)$.

⊗ **Key step:** Find NVS st $\sqrt{n}(\mu_n - \mu) * \gamma_\sigma$ converges & $\Phi = W_p$ is HDD

Smooth p -Wasserstein: Limit Distribution (Cont.)

Derivation: Take NVS as dual Sobolev space $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$

Smooth p -Wasserstein: Limit Distribution (Cont.)

Derivation: Take NVS as dual Sobolev space $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$

- ➊ **Smooth emp. process:** Previously seen convergence in $\dot{H}^{-1,p}(\gamma_\sigma)$

Smooth p -Wasserstein: Limit Distribution (Cont.)

Derivation: Take NVS as dual Sobolev space $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$

① **Smooth emp. process:** Previously seen convergence in $\dot{H}^{-1,p}(\gamma_\sigma)$

- ▶ Show $\sqrt{n}(\mu_n - \mu) * \gamma_\sigma \xrightarrow{w} G_\mu$ in $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$ via translation

Smooth p -Wasserstein: Limit Distribution (Cont.)

Derivation: Take NVS as dual Sobolev space $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$

① **Smooth emp. process:** Previously seen convergence in $\dot{H}^{-1,p}(\gamma_\sigma)$

- ▶ Show $\sqrt{n}(\mu_n - \mu) * \gamma_\sigma \xrightarrow{w} G_\mu$ in $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$ via translation

② **Func.:** For $D_\mu := \dot{H}^{-1,p}(\mu * \gamma_\sigma) \cap \{h = (\eta - \mu) * \gamma_\sigma, \eta \in \mathcal{P}_p(\mathbb{R}^d)\}$

Smooth p -Wasserstein: Limit Distribution (Cont.)

Derivation: Take NVS as dual Sobolev space $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$

① **Smooth emp. process:** Previously seen convergence in $\dot{H}^{-1,p}(\gamma_\sigma)$

- ▶ Show $\sqrt{n}(\mu_n - \mu) * \gamma_\sigma \xrightarrow{w} G_\mu$ in $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$ via translation

② **Func.:** For $D_\mu := \dot{H}^{-1,p}(\mu * \gamma_\sigma) \cap \{h = (\eta - \mu) * \gamma_\sigma, \eta \in \mathcal{P}_p(\mathbb{R}^d)\}$

- ▶ Set $\Phi : D_\mu \rightarrow \mathbb{R}$ as $\Phi(h) := W_p(\mu * \gamma_\sigma + h, \mu * \gamma_\sigma) = W_p^\sigma(\eta, \mu)$

Smooth p -Wasserstein: Limit Distribution (Cont.)

Derivation: Take NVS as dual Sobolev space $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$

① **Smooth emp. process:** Previously seen convergence in $\dot{H}^{-1,p}(\gamma_\sigma)$

- ▶ Show $\sqrt{n}(\mu_n - \mu) * \gamma_\sigma \xrightarrow{w} G_\mu$ in $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$ via translation

② **Func.:** For $D_\mu := \dot{H}^{-1,p}(\mu * \gamma_\sigma) \cap \{h = (\eta - \mu) * \gamma_\sigma, \eta \in \mathcal{P}_p(\mathbb{R}^d)\}$

- ▶ Set $\Phi : D_\mu \rightarrow \mathbb{R}$ as $\Phi(h) := W_p(\mu * \gamma_\sigma + h, \mu * \gamma_\sigma) = W_p^\sigma(\eta, \mu)$
- ▶ Observe that $W_p^\sigma(\mu_n, \mu) = \Phi((\mu_n - \mu) * \gamma_\sigma)$

Smooth p -Wasserstein: Limit Distribution (Cont.)

Derivation: Take NVS as dual Sobolev space $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$

① **Smooth emp. process:** Previously seen convergence in $\dot{H}^{-1,p}(\gamma_\sigma)$

- ▶ Show $\sqrt{n}(\mu_n - \mu) * \gamma_\sigma \xrightarrow{w} G_\mu$ in $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$ via translation

② **Func.:** For $D_\mu := \dot{H}^{-1,p}(\mu * \gamma_\sigma) \cap \{h = (\eta - \mu) * \gamma_\sigma, \eta \in \mathcal{P}_p(\mathbb{R}^d)\}$

- ▶ Set $\Phi : D_\mu \rightarrow \mathbb{R}$ as $\Phi(h) := W_p(\mu * \gamma_\sigma + h, \mu * \gamma_\sigma) = W_p^\sigma(\eta, \mu)$
- ▶ Observe that $W_p^\sigma(\mu_n, \mu) = \Phi((\mu_n - \mu) * \gamma_\sigma)$

③ **Derivative:** For locally Lipschitz Φ Hadamard DD = Gâteaux DD

Smooth p -Wasserstein: Limit Distribution (Cont.)

Derivation: Take NVS as dual Sobolev space $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$

① **Smooth emp. process:** Previously seen convergence in $\dot{H}^{-1,p}(\gamma_\sigma)$

- ▶ Show $\sqrt{n}(\mu_n - \mu) * \gamma_\sigma \xrightarrow{w} G_\mu$ in $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$ via translation

② **Func.:** For $D_\mu := \dot{H}^{-1,p}(\mu * \gamma_\sigma) \cap \{h = (\eta - \mu) * \gamma_\sigma, \eta \in \mathcal{P}_p(\mathbb{R}^d)\}$

- ▶ Set $\Phi : D_\mu \rightarrow \mathbb{R}$ as $\Phi(h) := W_p(\mu * \gamma_\sigma + h, \mu * \gamma_\sigma) = W_p^\sigma(\eta, \mu)$
- ▶ Observe that $W_p^\sigma(\mu_n, \mu) = \Phi((\mu_n - \mu) * \gamma_\sigma)$

③ **Derivative:** For locally Lipschitz Φ Hadamard DD = Gâteaux DD

- ▶ Comparison implies $\Phi(h) \lesssim \|h\|_{\dot{H}^{-1,p}(\mu * \gamma_\sigma)}$ & Derive Gâteaux DD

Smooth p -Wasserstein: Limit Distribution (Cont.)

Derivation: Take NVS as dual Sobolev space $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$

① **Smooth emp. process:** Previously seen convergence in $\dot{H}^{-1,p}(\gamma_\sigma)$

- ▶ Show $\sqrt{n}(\mu_n - \mu) * \gamma_\sigma \xrightarrow{w} G_\mu$ in $\dot{H}^{-1,p}(\mu * \gamma_\sigma)$ via translation

② **Func.:** For $D_\mu := \dot{H}^{-1,p}(\mu * \gamma_\sigma) \cap \{h = (\eta - \mu) * \gamma_\sigma, \eta \in \mathcal{P}_p(\mathbb{R}^d)\}$

- ▶ Set $\Phi : D_\mu \rightarrow \mathbb{R}$ as $\Phi(h) := W_p(\mu * \gamma_\sigma + h, \mu * \gamma_\sigma) = W_p^\sigma(\eta, \mu)$
- ▶ Observe that $W_p^\sigma(\mu_n, \mu) = \Phi((\mu_n - \mu) * \gamma_\sigma)$

③ **Derivative:** For locally Lipschitz Φ Hadamard DD = Gâteaux DD

- ▶ Comparison implies $\Phi(h) \lesssim \|h\|_{\dot{H}^{-1,p}(\mu * \gamma_\sigma)}$ & Derive Gâteaux DD

Theorem (Villani'03 for $p = 2$, ZG-Kato-Nietert-RiouxB'21)

For $\mu \in \mathcal{P}_p(\mathbb{R}^d)$, $h \in \dot{H}^{-1,p}(\mu)$ with $h \ll \mu$ and $\mu + h \in \mathcal{P}_p(\mathbb{R}^d)$, we have

$$\lim_{\lambda \downarrow 0} \frac{1}{\lambda} W_p(\mu + \lambda h, \mu) = \|h\|_{\dot{H}^{-1,p}(\mu)}.$$

Smooth p -Wasserstein: Limit Distribution (Cont. 2)

Theorem (ZG-Kato-Nietert-Rioux'21)

For $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ β -sub-Gaussian with $\beta < \sigma/\sqrt{2(p-1)}$, we have

$$\sqrt{n}W_p^\sigma(\mu_n, \mu) \xrightarrow{d} \sup_{\substack{g \in C_0^\infty: \\ \|g\|_{\dot{H}^{1,q}(\mu * \gamma_\sigma)} \leq 1}} G_\mu(g),$$

where $G_\mu = (G_\mu(g))_{g \in C_0^\infty}$ is a centered Gaussian process whose paths are linear and continuous wrt the Sobolev seminorm $\|\cdot\|_{\dot{H}^{1,q}(\mu * \gamma_\sigma)}$.

Smooth p -Wasserstein: Limit Distribution (Cont. 2)

Theorem (ZG-Kato-Nietert-Rioux'21)

For $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ β -sub-Gaussian with $\beta < \sigma / \sqrt{2(p-1)}$, we have

$$\sqrt{n}W_p^\sigma(\mu_n, \mu) \xrightarrow{d} \sup_{\substack{g \in C_0^\infty: \\ \|g\|_{\dot{H}^{1,q}(\mu * \gamma_\sigma)} \leq 1}} G_\mu(g),$$

where $G_\mu = (G_\mu(g))_{g \in C_0^\infty}$ is a centered Gaussian process whose paths are linear and continuous wrt the Sobolev seminorm $\|\cdot\|_{\dot{H}^{1,q}(\mu * \gamma_\sigma)}$.

Comments:

- More results: One- and two-sample cases under null and alternative

Smooth p -Wasserstein: Limit Distribution (Cont. 2)

Theorem (ZG-Kato-Nietert-Rioux'21)

For $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ β -sub-Gaussian with $\beta < \sigma/\sqrt{2(p-1)}$, we have

$$\sqrt{n}W_p^\sigma(\mu_n, \mu) \xrightarrow{d} \sup_{\substack{g \in C_0^\infty: \\ \|g\|_{\dot{H}^{1,q}(\mu * \gamma_\sigma)} \leq 1}} G_\mu(g),$$

where $G_\mu = (G_\mu(g))_{g \in C_0^\infty}$ is a centered Gaussian process whose paths are linear and continuous wrt the Sobolev seminorm $\|\cdot\|_{\dot{H}^{1,q}(\mu * \gamma_\sigma)}$.

Comments:

- **More results:** One- and two-sample cases under null and alternative
- **Asymptotic normality:** For both results under alternative (HDD linear)

Smooth p -Wasserstein: Limit Distribution (Cont. 2)

Theorem (ZG-Kato-Nietert-Rioux'21)

For $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ β -sub-Gaussian with $\beta < \sigma/\sqrt{2(p-1)}$, we have

$$\sqrt{n}W_p^\sigma(\mu_n, \mu) \xrightarrow{d} \sup_{\substack{g \in C_0^\infty: \\ \|g\|_{\dot{H}^{1,q}(\mu * \gamma_\sigma)} \leq 1}} G_\mu(g),$$

where $G_\mu = (G_\mu(g))_{g \in C_0^\infty}$ is a centered Gaussian process whose paths are linear and continuous wrt the Sobolev seminorm $\|\cdot\|_{\dot{H}^{1,q}(\mu * \gamma_\sigma)}$.

Comments:

- **More results:** One- and two-sample cases under null and alternative
- **Asymptotic normality:** For both results under alternative (HDD linear)
- **Bootstrap:** Prove consistency of the nonparametric bootstrap

Smooth p -Wasserstein: Limit Distribution (Cont. 2)

Theorem (ZG-Kato-Nietert-Rioux'21)

For $\mu \in \mathcal{P}_p(\mathbb{R}^d)$ β -sub-Gaussian with $\beta < \sigma/\sqrt{2(p-1)}$, we have

$$\sqrt{n}W_p^\sigma(\mu_n, \mu) \xrightarrow{d} \sup_{\substack{g \in C_0^\infty: \\ \|g\|_{\dot{H}^{1,q}(\mu * \gamma_\sigma)} \leq 1}} G_\mu(g),$$

where $G_\mu = (G_\mu(g))_{g \in C_0^\infty}$ is a centered Gaussian process whose paths are linear and continuous wrt the Sobolev seminorm $\|\cdot\|_{\dot{H}^{1,q}(\mu * \gamma_\sigma)}$.

Comments:

- **More results:** One- and two-sample cases under null and alternative
- **Asymptotic normality:** For both results under alternative (HDD linear)
- **Bootstrap:** Prove consistency of the nonparametric bootstrap

All structural and statistical properties of W_1^σ transfer to W_p^σ

Summary and Concluding Remarks

- **Classic W_p :** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ w/ rich structure & many applications

Summary and Concluding Remarks

- **Classic W_p :** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ w/ rich structure & many applications
 - ▶ Generative modeling, testing, nonparametric mixture estimation, etc.

Summary and Concluding Remarks

- **Classic W_p :** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ w/ rich structure & many applications
 - ▶ Generative modeling, testing, nonparametric mixture estimation, etc.
 - ▶ Empirical approximation is slow $n^{-1/d}$

Summary and Concluding Remarks

- **Classic W_p :** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ w/ rich structure & many applications
 - ▶ Generative modeling, testing, nonparametric mixture estimation, etc.
 - ▶ Empirical approximation is slow $n^{-1/d}$
- **Smooth W_p^σ :** Convolve distributions w/ Gaussian kernel

Summary and Concluding Remarks

- **Classic W_p :** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ w/ rich structure & many applications
 - ▶ Generative modeling, testing, nonparametric mixture estimation, etc.
 - ▶ Empirical approximation is slow $n^{-1/d}$
- **Smooth W_p^σ :** Convolve distributions w/ Gaussian kernel
 - ▶ Inherits structure of W_p (metric, topology, duality)

Summary and Concluding Remarks

- **Classic W_p :** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ w/ rich structure & many applications
 - ▶ Generative modeling, testing, nonparametric mixture estimation, etc.
 - ▶ Empirical approximation is slow $n^{-1/d}$
- **Smooth W_p^σ :** Convolve distributions w/ Gaussian kernel
 - ▶ Inherits structure of W_p (metric, topology, duality)
 - ▶ Well-behaved function of smoothing parameter & recovers W_p in limit

Summary and Concluding Remarks

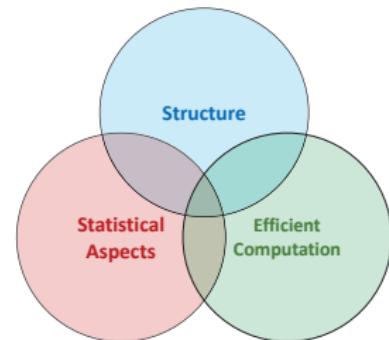
- **Classic W_p :** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ w/ rich structure & many applications
 - ▶ Generative modeling, testing, nonparametric mixture estimation, etc.
 - ▶ Empirical approximation is slow $n^{-1/d}$
- **Smooth W_p^σ :** Convolve distributions w/ Gaussian kernel
 - ▶ Inherits structure of W_p (metric, topology, duality)
 - ▶ Well-behaved function of smoothing parameter & recovers W_p in limit
 - ▶ Fast $n^{-1/2}$ empirical convergence in all dimensions

Summary and Concluding Remarks

- **Classic W_p :** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ w/ rich structure & many applications
 - ▶ Generative modeling, testing, nonparametric mixture estimation, etc.
 - ▶ Empirical approximation is slow $n^{-1/d}$
- **Smooth W_p^σ :** Convolve distributions w/ Gaussian kernel
 - ▶ Inherits structure of W_p (metric, topology, duality)
 - ▶ Well-behaved function of smoothing parameter & recovers W_p in limit
 - ▶ Fast $n^{-1/2}$ empirical convergence in all dimensions
 - ▶ Comprehensive limit distribution theory for empirical $\sqrt{n}W_p^\sigma$ in all dim.

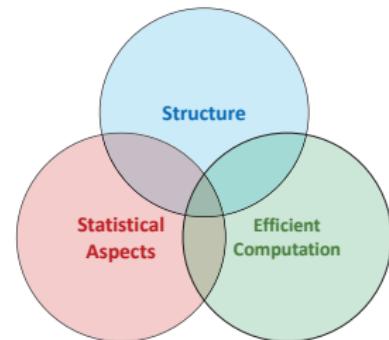
Summary and Concluding Remarks

- **Classic W_p :** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ w/ rich structure & many applications
 - ▶ Generative modeling, testing, nonparametric mixture estimation, etc.
 - ▶ Empirical approximation is slow $n^{-1/d}$
- **Smooth W_p^σ :** Convolve distributions w/ Gaussian kernel
 - ▶ Inherits structure of W_p (metric, topology, duality)
 - ▶ Well-behaved function of smoothing parameter & recovers W_p in limit
 - ▶ Fast $n^{-1/2}$ empirical convergence in all dimensions
 - ▶ Comprehensive limit distribution theory for empirical $\sqrt{n}W_p^\sigma$ in all dim.



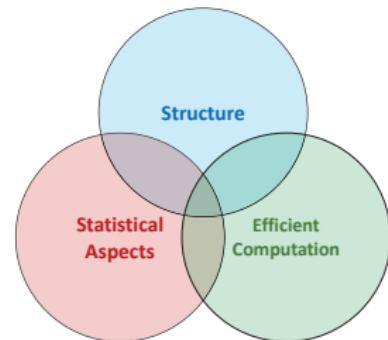
Summary and Concluding Remarks

- **Classic W_p :** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ w/ rich structure & many applications
 - ▶ Generative modeling, testing, nonparametric mixture estimation, etc.
 - ▶ Empirical approximation is slow $n^{-1/d}$
- **Smooth W_p^σ :** Convolve distributions w/ Gaussian kernel
 - ▶ Inherits structure of W_p (metric, topology, duality)
 - ▶ Well-behaved function of smoothing parameter & recovers W_p in limit
 - ▶ Fast $n^{-1/2}$ empirical convergence in all dimensions
 - ▶ Comprehensive limit distribution theory for empirical $\sqrt{n}W_p^\sigma$ in all dim.
- **Key open question:** Efficient computation



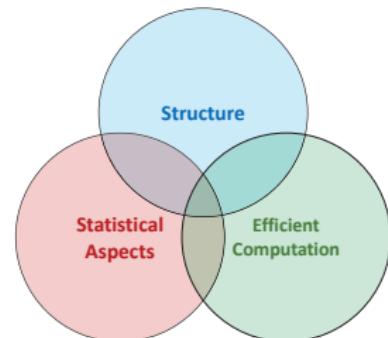
Summary and Concluding Remarks

- **Classic W_p :** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ w/ rich structure & many applications
 - ▶ Generative modeling, testing, nonparametric mixture estimation, etc.
 - ▶ Empirical approximation is slow $n^{-1/d}$
- **Smooth W_p^σ :** Convolve distributions w/ Gaussian kernel
 - ▶ Inherits structure of W_p (metric, topology, duality)
 - ▶ Well-behaved function of smoothing parameter & recovers W_p in limit
 - ▶ Fast $n^{-1/2}$ empirical convergence in all dimensions
 - ▶ Comprehensive limit distribution theory for empirical $\sqrt{n}W_p^\sigma$ in all dim.
- **Key open question:** Efficient computation
 - ▶ Optimal transport btw Gaussian mixtures?



Summary and Concluding Remarks

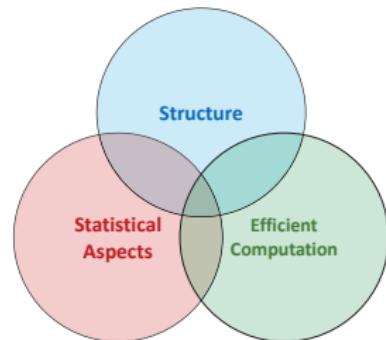
- **Classic W_p :** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ w/ rich structure & many applications
 - ▶ Generative modeling, testing, nonparametric mixture estimation, etc.
 - ▶ Empirical approximation is slow $n^{-1/d}$
- **Smooth W_p^σ :** Convolve distributions w/ Gaussian kernel
 - ▶ Inherits structure of W_p (metric, topology, duality)
 - ▶ Well-behaved function of smoothing parameter & recovers W_p in limit
 - ▶ Fast $n^{-1/2}$ empirical convergence in all dimensions
 - ▶ Comprehensive limit distribution theory for empirical $\sqrt{n}W_p^\sigma$ in all dim.
- **Key open question:** Efficient computation
 - ▶ Optimal transport btw Gaussian mixtures?
 - ▶ Compactly supported smoothing kernels?



Summary and Concluding Remarks

- **Classic W_p :** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ w/ rich structure & many applications
 - ▶ Generative modeling, testing, nonparametric mixture estimation, etc.
 - ▶ Empirical approximation is slow $n^{-1/d}$
- **Smooth W_p^σ :** Convolve distributions w/ Gaussian kernel
 - ▶ Inherits structure of W_p (metric, topology, duality)
 - ▶ Well-behaved function of smoothing parameter & recovers W_p in limit
 - ▶ Fast $n^{-1/2}$ empirical convergence in all dimensions
 - ▶ Comprehensive limit distribution theory for empirical $\sqrt{n}W_p^\sigma$ in all dim.
- **Key open question:** Efficient computation
 - ▶ Optimal transport btw Gaussian mixtures?
 - ▶ Compactly supported smoothing kernels?

Thank you!



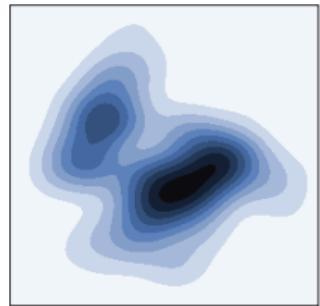
Implicit Generative Models: Generalization

Goal: Solve $\text{OPT} := \inf_{\theta} W_1(\mu, \nu_{\theta})$ exactly (find θ^*)

Implicit Generative Models: Generalization

Goal: Solve $\text{OPT} := \inf_{\theta} W_1(\mu, \nu_{\theta})$ exactly (find θ^*)

Estimation: We don't have μ but data

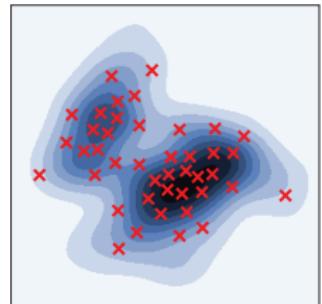


Implicit Generative Models: Generalization

Goal: Solve $\text{OPT} := \inf_{\theta} W_1(\mu, \nu_{\theta})$ exactly (find θ^*)

Estimation: We don't have μ but data

- $\{X_i\}_{i=1}^n$ are i.i.d. samples from $\mu \in \mathcal{P}(\mathbb{R}^d)$

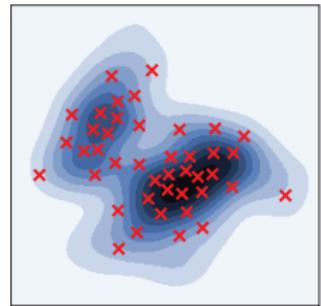


Implicit Generative Models: Generalization

Goal: Solve $\text{OPT} := \inf_{\theta} W_1(\mu, \nu_{\theta})$ exactly (find θ^*)

Estimation: We don't have μ but data

- $\{X_i\}_{i=1}^n$ are i.i.d. samples from $\mu \in \mathcal{P}(\mathbb{R}^d)$
- Empirical distribution $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$

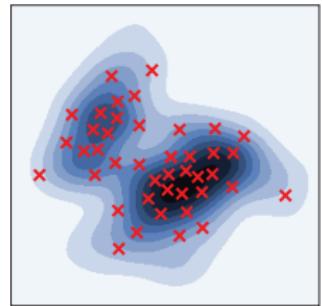


Implicit Generative Models: Generalization

Goal: Solve $\text{OPT} := \inf_{\theta} W_1(\mu, \nu_{\theta})$ exactly (find θ^*)

Estimation: We don't have μ but data

- $\{X_i\}_{i=1}^n$ are i.i.d. samples from $\mu \in \mathcal{P}(\mathbb{R}^d)$
 - Empirical distribution $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$
- ⇒ Inherently we work with $W_1(\mu_n, \nu_{\theta})$

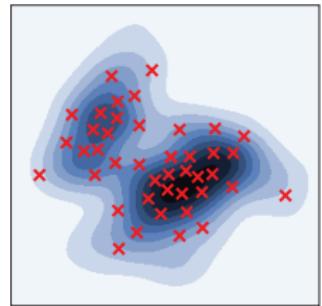


Implicit Generative Models: Generalization

Goal: Solve $\text{OPT} := \inf_{\theta} W_1(\mu, \nu_{\theta})$ exactly (find θ^*)

Estimation: We don't have μ but data

- $\{X_i\}_{i=1}^n$ are i.i.d. samples from $\mu \in \mathcal{P}(\mathbb{R}^d)$
 - Empirical distribution $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$
- ⇒ Inherently we work with $W_1(\mu_n, \nu_{\theta})$



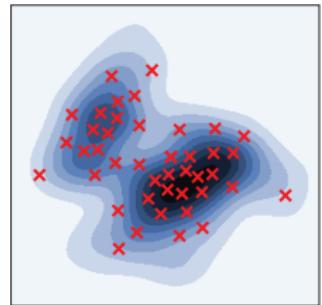
Optimization: Can solve $\inf_{\theta} W_1(\mu_n, \nu_{\theta})$ approximately

Implicit Generative Models: Generalization

Goal: Solve $\text{OPT} := \inf_{\theta} W_1(\mu, \nu_{\theta})$ exactly (find θ^*)

Estimation: We don't have μ but data

- $\{X_i\}_{i=1}^n$ are i.i.d. samples from $\mu \in \mathcal{P}(\mathbb{R}^d)$
 - Empirical distribution $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$
- ⇒ Inherently we work with $W_1(\mu_n, \nu_{\theta})$



Optimization: Can solve $\inf_{\theta} W_1(\mu_n, \nu_{\theta})$ approximately

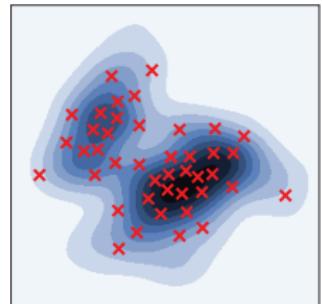
$$\text{Find } \hat{\theta}_n \text{ s.t. } W_1(\mu_n, \nu_{\hat{\theta}_n}) \leq \inf_{\theta} W_1(\mu_n, \nu_{\theta}) + \epsilon$$

Implicit Generative Models: Generalization

Goal: Solve $\text{OPT} := \inf_{\theta} W_1(\mu, \nu_{\theta})$ exactly (find θ^*)

Estimation: We don't have μ but data

- $\{X_i\}_{i=1}^n$ are i.i.d. samples from $\mu \in \mathcal{P}(\mathbb{R}^d)$
 - Empirical distribution $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$
- ⇒ Inherently we work with $W_1(\mu_n, \nu_{\theta})$



Optimization: Can solve $\inf_{\theta} W_1(\mu_n, \nu_{\theta})$ approximately

$$\text{Find } \hat{\theta}_n \text{ s.t. } W_1(\mu_n, \nu_{\hat{\theta}_n}) \leq \inf_{\theta} W_1(\mu_n, \nu_{\theta}) + \epsilon$$

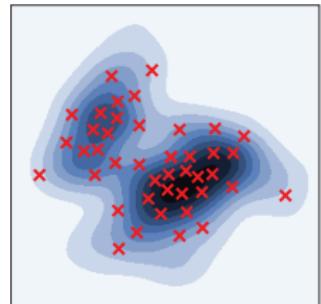
Generalization: $W_1(\mu, \nu_{\hat{\theta}_n}) - \text{OPT}$

Implicit Generative Models: Generalization

Goal: Solve $\text{OPT} := \inf_{\theta} W_1(\mu, \nu_{\theta})$ exactly (find θ^*)

Estimation: We don't have μ but data

- $\{X_i\}_{i=1}^n$ are i.i.d. samples from $\mu \in \mathcal{P}(\mathbb{R}^d)$
 - Empirical distribution $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$
- ⇒ Inherently we work with $W_1(\mu_n, \nu_{\theta})$



Optimization: Can solve $\inf_{\theta} W_1(\mu_n, \nu_{\theta})$ approximately

$$\text{Find } \hat{\theta}_n \text{ s.t. } W_1(\mu_n, \nu_{\hat{\theta}_n}) \leq \inf_{\theta} W_1(\mu_n, \nu_{\theta}) + \epsilon$$

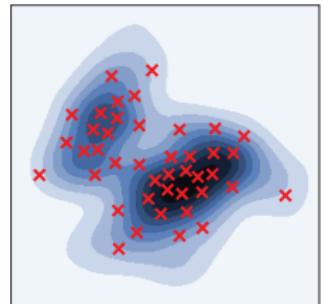
Generalization: $W_1(\mu, \nu_{\hat{\theta}_n}) - \text{OPT} \leq 2W_1(\mu_n, \mu) + \epsilon$

Implicit Generative Models: Generalization

Goal: Solve $\text{OPT} := \inf_{\theta} W_1(\mu, \nu_{\theta})$ exactly (find θ^*)

Estimation: We don't have μ but data

- $\{X_i\}_{i=1}^n$ are i.i.d. samples from $\mu \in \mathcal{P}(\mathbb{R}^d)$
 - Empirical distribution $\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$
- ⇒ Inherently we work with $W_1(\mu_n, \nu_{\theta})$



Optimization: Can solve $\inf_{\theta} W_1(\mu_n, \nu_{\theta})$ approximately

$$\text{Find } \hat{\theta}_n \text{ s.t. } W_1(\mu_n, \nu_{\hat{\theta}_n}) \leq \inf_{\theta} W_1(\mu_n, \nu_{\theta}) + \epsilon$$

Generalization: $W_1(\mu, \nu_{\hat{\theta}_n}) - \text{OPT} \leq 2W_1(\mu_n, \mu) + \epsilon$

⇒ Boils down to empirical approximation question under W_1

Directional Derivative: Proof Outline

Objective: Show that $\lim_{\lambda \downarrow 0} \frac{1}{\lambda} W_p(\mu + \lambda h, \mu) = \|h\|_{\dot{H}^{-1,p}(\mu)}$

Directional Derivative: Proof Outline

Objective: Show that $\lim_{\lambda \downarrow 0} \frac{1}{\lambda} W_p(\mu + \lambda h, \mu) = \|h\|_{\dot{H}^{-1,p}(\mu)}$

- ① For $\lambda \ll 1$ dynamic formulation simplifies:

$$W_p(\mu + \lambda h, \mu) = \inf_{\substack{\mu_t, v_t: \\ \partial_t \mu_t = -\nabla \cdot (\mu_t v_t) \\ \mu_0 = \mu, \mu_1 = \mu + \lambda h}} \int_0^1 \|v_t\|_{L^p(\mu_t)} dt$$

Directional Derivative: Proof Outline

Objective: Show that $\lim_{\lambda \downarrow 0} \frac{1}{\lambda} W_p(\mu + \lambda h, \mu) = \|h\|_{\dot{H}^{-1,p}(\mu)}$

- ① For $\lambda \ll 1$ dynamic formulation simplifies:

$$W_p(\mu + \lambda h, \mu) = \inf_{\substack{\mu_t, v_t: \\ \partial_t \mu_t = -\nabla \cdot (\mu_t v_t) \\ \mu_0 = \mu, \mu_1 = \mu + \lambda h}} \int_0^1 \|v_t\|_{L^p(\mu_t)} dt \stackrel{\substack{\mu_t \approx \mu + t\lambda h \\ v_t \approx v}}{=} \quad$$

Directional Derivative: Proof Outline

Objective: Show that $\lim_{\lambda \downarrow 0} \frac{1}{\lambda} W_p(\mu + \lambda h, \mu) = \|h\|_{\dot{H}^{-1,p}(\mu)}$

- ① For $\lambda \ll 1$ dynamic formulation simplifies:

$$W_p(\mu + \lambda h, \mu) = \inf_{\substack{\mu_t, v_t: \\ \partial_t \mu_t = -\nabla \cdot (\mu_t v_t) \\ \mu_0 = \mu, \mu_1 = \mu + \lambda h}} \int_0^1 \|v_t\|_{L^p(\mu_t)} dt \stackrel{\substack{\mu_t \approx \mu + t\lambda h \\ v_t \approx v}}{=} \inf_{\substack{v: \\ \lambda h = -\nabla \cdot (\mu v)}} \|v\|_{L^p(\mu)} + o(\lambda)$$

Directional Derivative: Proof Outline

Objective: Show that $\lim_{\lambda \downarrow 0} \frac{1}{\lambda} W_p(\mu + \lambda h, \mu) = \|h\|_{\dot{H}^{-1,p}(\mu)}$

- ① For $\lambda \ll 1$ dynamic formulation simplifies:

$$W_p(\mu + \lambda h, \mu) = \inf_{\substack{\mu_t, v_t: \\ \partial_t \mu_t = -\nabla \cdot (\mu_t v_t) \\ \mu_0 = \mu, \mu_1 = \mu + \lambda h}} \int_0^1 \|v_t\|_{L^p(\mu_t)} dt \stackrel{\substack{\mu_t \approx \mu + t\lambda h \\ v_t \approx v}}{=} \inf_{\substack{v: \\ \lambda h = -\nabla \cdot (\mu v)}} \|v\|_{L^p(\mu)} + o(\lambda)$$

- ② Let u be a weak solution to $-\Delta u \rho = \mu - \nu$:

Directional Derivative: Proof Outline

Objective: Show that $\lim_{\lambda \downarrow 0} \frac{1}{\lambda} W_p(\mu + \lambda h, \mu) = \|h\|_{\dot{H}^{-1,p}(\mu)}$

- ① For $\lambda \ll 1$ dynamic formulation simplifies:

$$W_p(\mu + \lambda h, \mu) = \inf_{\substack{\mu_t, v_t: \\ \partial_t \mu_t = -\nabla \cdot (\mu_t v_t) \\ \mu_0 = \mu, \mu_1 = \mu + \lambda h}} \int_0^1 \|v_t\|_{L^p(\mu_t)} dt \stackrel{\substack{\mu_t \approx \mu + t\lambda h \\ v_t \approx v}}{=} \inf_{\substack{v: \\ \lambda h = -\nabla \cdot (\mu v)}} \|v\|_{L^p(\mu)} + o(\lambda)$$

- ② Let u be a weak solution to $-\Delta u \rho = \mu - \nu$:

$$\int f d(\mu - \nu) = - \int f \Delta u d\rho$$

Directional Derivative: Proof Outline

Objective: Show that $\lim_{\lambda \downarrow 0} \frac{1}{\lambda} W_p(\mu + \lambda h, \mu) = \|h\|_{\dot{H}^{-1,p}(\mu)}$

- ① For $\lambda \ll 1$ dynamic formulation simplifies:

$$W_p(\mu + \lambda h, \mu) = \inf_{\substack{\mu_t, v_t: \\ \partial_t \mu_t = -\nabla \cdot (\mu_t v_t) \\ \mu_0 = \mu, \mu_1 = \mu + \lambda h}} \int_0^1 \|v_t\|_{L^p(\mu_t)} dt \stackrel{\substack{\mu_t \approx \mu + t\lambda h \\ v_t \approx v}}{=} \inf_{\substack{v: \\ \lambda h = -\nabla \cdot (\mu v)}} \|v\|_{L^p(\mu)} + o(\lambda)$$

- ② Let u be a weak solution to $-\Delta u \rho = \mu - \nu$:

$$\int f d(\mu - \nu) = - \int f \Delta u d\rho = \int \nabla f \cdot \nabla u d\rho$$

Directional Derivative: Proof Outline

Objective: Show that $\lim_{\lambda \downarrow 0} \frac{1}{\lambda} W_p(\mu + \lambda h, \mu) = \|h\|_{\dot{H}^{-1,p}(\mu)}$

- ① For $\lambda \ll 1$ dynamic formulation simplifies:

$$W_p(\mu + \lambda h, \mu) = \inf_{\substack{\mu_t, v_t: \\ \partial_t \mu_t = -\nabla \cdot (\mu_t v_t) \\ \mu_0 = \mu, \mu_1 = \mu + \lambda h}} \int_0^1 \|v_t\|_{L^p(\mu_t)} dt \stackrel{\substack{\mu_t \approx \mu + t\lambda h \\ v_t \approx v}}{=} \inf_{\substack{v: \\ \lambda h = -\nabla \cdot (\mu v)}} \|v\|_{L^p(\mu)} + o(\lambda)$$

- ② Let u be a weak solution to $-\Delta u \rho = \mu - \nu$:

$$\int f d(\mu - \nu) = - \int f \Delta u d\rho = \int \nabla f \cdot \nabla u d\rho \leq \|\nabla f\|_{L^q(\rho)} \|\nabla u\|_{L^p(\rho)}$$

Directional Derivative: Proof Outline

Objective: Show that $\lim_{\lambda \downarrow 0} \frac{1}{\lambda} W_p(\mu + \lambda h, \mu) = \|h\|_{\dot{H}^{-1,p}(\mu)}$

- ① For $\lambda \ll 1$ dynamic formulation simplifies:

$$W_p(\mu + \lambda h, \mu) = \inf_{\substack{\mu_t, v_t: \\ \partial_t \mu_t = -\nabla \cdot (\mu_t v_t) \\ \mu_0 = \mu, \mu_1 = \mu + \lambda h}} \int_0^1 \|v_t\|_{L^p(\mu_t)} dt \stackrel{\substack{\mu_t \approx \mu + t\lambda h \\ v_t \approx v}}{=} \inf_{\substack{v: \\ \lambda h = -\nabla \cdot (\mu v)}} \|v\|_{L^p(\mu)} + o(\lambda)$$

- ② Let u be a weak solution to $-\Delta u \rho = \mu - \nu$:

$$\int f d(\mu - \nu) = - \int f \Delta u d\rho = \int \nabla f \cdot \nabla u d\rho \leq \|\nabla f\|_{L^q(\rho)} \|\nabla u\|_{L^p(\rho)}$$

$$\xrightarrow{v = \nabla u} \|\mu - \nu\|_{\dot{H}^{-1,p}(\rho)} = \sup_{\|\nabla f\|_{L^q(\rho)} \leq 1} \int f d(\mu - \nu) = \inf_{\substack{v: \\ \mu - \nu = -\nabla \cdot (\rho v)}} \|v\|_{L^p(\rho)}$$

Directional Derivative: Proof Outline

Objective: Show that $\lim_{\lambda \downarrow 0} \frac{1}{\lambda} W_p(\mu + \lambda h, \mu) = \|h\|_{\dot{H}^{-1,p}(\mu)}$

- ① For $\lambda \ll 1$ dynamic formulation simplifies:

$$W_p(\mu + \lambda h, \mu) = \inf_{\substack{\mu_t, v_t: \\ \partial_t \mu_t = -\nabla \cdot (\mu_t v_t) \\ \mu_0 = \mu, \mu_1 = \mu + \lambda h}} \int_0^1 \|v_t\|_{L^p(\mu_t)} dt \stackrel{\begin{array}{l} \mu_t \approx \mu + t \lambda h \\ v_t \approx v \end{array}}{=} \inf_{\substack{v: \\ \lambda h = -\nabla \cdot (\mu v)}} \|v\|_{L^p(\mu)} + o(\lambda)$$

- ② Let u be a weak solution to $-\Delta u \rho = \mu - \nu$:

$$\int f d(\mu - \nu) = - \int f \Delta u d\rho = \int \nabla f \cdot \nabla u d\rho \leq \|\nabla f\|_{L^q(\rho)} \|\nabla u\|_{L^p(\rho)}$$

$$\xrightarrow{v = \nabla u} \|\mu - \nu\|_{\dot{H}^{-1,p}(\rho)} = \sup_{\|\nabla f\|_{L^q(\rho)} \leq 1} \int f d(\mu - \nu) = \inf_{\substack{v: \\ \mu - \nu = -\nabla \cdot (\rho v)}} \|v\|_{L^p(\rho)}$$

Directional Derivative: Proof Outline

Objective: Show that $\lim_{\lambda \downarrow 0} \frac{1}{\lambda} W_p(\mu + \lambda h, \mu) = \|h\|_{\dot{H}^{-1,p}(\mu)}$

- ① For $\lambda \ll 1$ dynamic formulation simplifies:

$$W_p(\mu + \lambda h, \mu) = \inf_{\substack{\mu_t, v_t: \\ \partial_t \mu_t = -\nabla \cdot (\mu_t v_t) \\ \mu_0 = \mu, \mu_1 = \mu + \lambda h}} \int_0^1 \|v_t\|_{L^p(\mu_t)} dt \stackrel{\begin{array}{l} \mu_t \approx \mu + t \lambda h \\ v_t \approx v \end{array}}{=} \inf_{\substack{v: \\ \lambda h = -\nabla \cdot (\mu v)}} \|v\|_{L^p(\mu)} + o(\lambda)$$

- ② Let u be a weak solution to $-\Delta u \rho = \mu - \nu$:

$$\int f d(\mu - \nu) = - \int f \Delta u d\rho = \int \nabla f \cdot \nabla u d\rho \leq \|\nabla f\|_{L^q(\rho)} \|\nabla u\|_{L^p(\rho)}$$

$$\xrightarrow{v = \nabla u} \|\mu - \nu\|_{\dot{H}^{-1,p}(\rho)} = \sup_{\|\nabla f\|_{L^q(\rho)} \leq 1} \int f d(\mu - \nu) = \inf_{\substack{v: \\ \mu - \nu = -\nabla \cdot (\rho v)}} \|v\|_{L^p(\rho)}$$

- ③ Conclude:

$$\lim_{\lambda \downarrow 0} \frac{1}{\lambda} W_p(\mu + \lambda h, \mu)$$

Directional Derivative: Proof Outline

Objective: Show that $\lim_{\lambda \downarrow 0} \frac{1}{\lambda} W_p(\mu + \lambda h, \mu) = \|h\|_{\dot{H}^{-1,p}(\mu)}$

- ① For $\lambda \ll 1$ dynamic formulation simplifies:

$$W_p(\mu + \lambda h, \mu) = \inf_{\substack{\mu_t, v_t: \\ \partial_t \mu_t = -\nabla \cdot (\mu_t v_t) \\ \mu_0 = \mu, \mu_1 = \mu + \lambda h}} \int_0^1 \|v_t\|_{L^p(\mu_t)} dt \stackrel{\begin{array}{l} \mu_t \approx \mu + t \lambda h \\ v_t \approx v \end{array}}{=} \inf_{\substack{v: \\ \lambda h = -\nabla \cdot (\mu v)}} \|v\|_{L^p(\mu)} + o(\lambda)$$

- ② Let u be a weak solution to $-\Delta u \rho = \mu - \nu$:

$$\int f d(\mu - \nu) = - \int f \Delta u d\rho = \int \nabla f \cdot \nabla u d\rho \leq \|\nabla f\|_{L^q(\rho)} \|\nabla u\|_{L^p(\rho)}$$

$$\xrightarrow{v = \nabla u} \|\mu - \nu\|_{\dot{H}^{-1,p}(\rho)} = \sup_{\|\nabla f\|_{L^q(\rho)} \leq 1} \int f d(\mu - \nu) = \inf_{\substack{v: \\ \mu - \nu = -\nabla \cdot (\rho v)}} \|v\|_{L^p(\rho)}$$

- ③ Conclude:

$$\lim_{\lambda \downarrow 0} \frac{1}{\lambda} W_p(\mu + \lambda h, \mu) = \lim_{\lambda \downarrow 0} \frac{1}{\lambda} \left(\|\lambda h\|_{\dot{H}^{-1,p}(\mu)} + o(\lambda) \right)$$

Directional Derivative: Proof Outline

Objective: Show that $\lim_{\lambda \downarrow 0} \frac{1}{\lambda} W_p(\mu + \lambda h, \mu) = \|h\|_{\dot{H}^{-1,p}(\mu)}$

- ① For $\lambda \ll 1$ dynamic formulation simplifies:

$$W_p(\mu + \lambda h, \mu) = \inf_{\substack{\mu_t, v_t: \\ \partial_t \mu_t = -\nabla \cdot (\mu_t v_t) \\ \mu_0 = \mu, \mu_1 = \mu + \lambda h}} \int_0^1 \|v_t\|_{L^p(\mu_t)} dt \stackrel{\substack{\mu_t \approx \mu + t\lambda h \\ v_t \approx v}}{=} \inf_{\substack{v: \\ \lambda h = -\nabla \cdot (\mu v)}} \|v\|_{L^p(\mu)} + o(\lambda)$$

- ② Let u be a weak solution to $-\Delta u \rho = \mu - \nu$:

$$\int f d(\mu - \nu) = - \int f \Delta u d\rho = \int \nabla f \cdot \nabla u d\rho \leq \|\nabla f\|_{L^q(\rho)} \|\nabla u\|_{L^p(\rho)}$$

$$\xrightarrow{v = \nabla u} \|\mu - \nu\|_{\dot{H}^{-1,p}(\rho)} = \sup_{\|\nabla f\|_{L^q(\rho)} \leq 1} \int f d(\mu - \nu) = \inf_{\substack{v: \\ \mu - \nu = -\nabla \cdot (\rho v)}} \|v\|_{L^p(\rho)}$$

- ③ Conclude:

$$\lim_{\lambda \downarrow 0} \frac{1}{\lambda} W_p(\mu + \lambda h, \mu) = \lim_{\lambda \downarrow 0} \frac{1}{\lambda} \left(\|\lambda h\|_{\dot{H}^{-1,p}(\mu)} + o(\lambda) \right) = \|h\|_{\dot{H}^{-1,p}(\mu)}$$