Scaling Wasserstein Distances to High Dimensions via Smoothing

Ziv Goldfeld

Cornell University

Implicit (Latent Variable) Generative Models

Implicit (Latent Variable) Generative Models

Goal: Learn a model $Q_\theta \approx P \in \mathcal{P}(\mathbb{R}^d)$ to approximate data distribution

Implicit (Latent Variable) Generative Models

Goal: Learn a model $Q_\theta \approx P \in \mathcal{P}(\mathbb{R}^d)$ to approximate data distribution

Method: Complicated transformation of a simple latent variable

Implicit (Latent Variable) Generative Models

Goal: Learn a model $Q_\theta \approx P \in \mathcal{P}(\mathbb{R}^d)$ to approximate data distribution

Method: Complicated transformation of a simple latent variable

- Latent variable $Z \sim Q_Z \in \mathcal{P}(\mathbb{R}^p)$, $p \ll d$

Implicit (Latent Variable) Generative Models

Goal: Learn a model $Q_\theta \approx P \in \mathcal{P}(\mathbb{R}^d)$ to approximate data distribution

Method: Complicated transformation of a simple latent variable

- Latent variable $Z \sim Q_Z \in \mathcal{P}(\mathbb{R}^p)$, $p \ll d$
- Expand Z to $\mathbb{R}^d$ space via (random) transformation $Q^{(\theta)}_{X|Z}$

Implicit (Latent Variable) Generative Models

Goal: Learn a model $Q_\theta \approx P \in \mathcal{P}(\mathbb{R}^d)$ to approximate data distribution

Method: Complicated transformation of a simple latent variable

- Latent variable $Z \sim Q_Z \in \mathcal{P}(\mathbb{R}^p)$, $p \ll d$
- Expand Z to $\mathbb{R}^d$ space via (random) transformation $Q^{(\theta)}_{X|Z}$

$\Longrightarrow$ **Generative model:** $Q_\theta(\cdot) := \int_{\mathbb{R}^p} Q^{(\theta)}_{X|Z}(\cdot|z)\,\mathrm{d}Q_Z(z)$

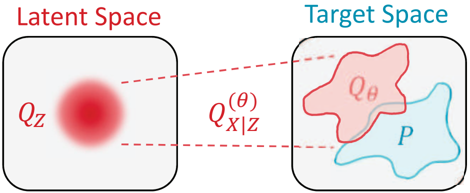
Implicit (Latent Variable) Generative Models

Goal: Learn a model $Q_\theta \approx P \in \mathcal{P}(\mathbb{R}^d)$ to approximate data distribution

Method: Complicated transformation of a simple latent variable

- Latent variable $Z \sim Q_Z \in \mathcal{P}(\mathbb{R}^p)$, $p \ll d$
- Expand Z to $\mathbb{R}^d$ space via (random) transformation $Q^{(\theta)}_{X|Z}$

$\Longrightarrow$ **Generative model:** $Q_\theta(\cdot) := \int_{\mathbb{R}^p} Q^{(\theta)}_{X|Z}(\cdot|z)\,\mathrm{d}Q_Z(z)$



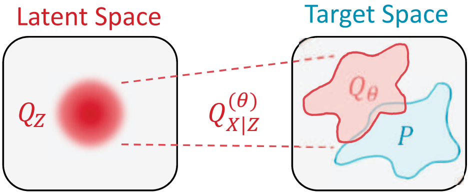
Implicit (Latent Variable) Generative Models

Goal: Learn a model $Q_\theta \approx P \in \mathcal{P}(\mathbb{R}^d)$ to approximate data distribution

Method: Complicated transformation of a simple latent variable

- Latent variable $Z \sim Q_Z \in \mathcal{P}(\mathbb{R}^p)$, $p \ll d$
- Expand Z to $\mathbb{R}^d$ space via (random) transformation $Q^{(\theta)}_{X|Z}$

$\Longrightarrow$ **Generative model:** $Q_\theta(\cdot) := \int_{\mathbb{R}^p} Q^{(\theta)}_{X|Z}(\cdot|z)\, \mathrm{d}Q_Z(z)$



Minimum Distance Estimation: Solve $\boxed{\theta^\star \in \underset{\theta}{\operatorname{argmin}}\, \delta(P, Q_\theta)}$

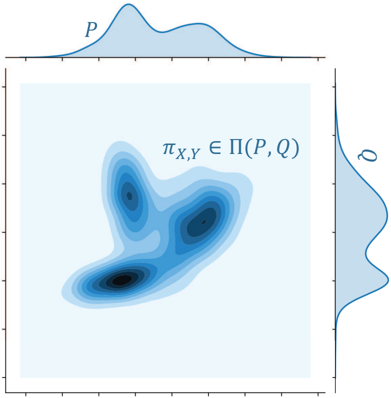
The p-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite pth moments)

The p-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite pth moments)

- **Coupling:** $\Pi(P,Q) = \left\{\pi_{X,Y} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \ : \ \pi_X = P \ \& \ \pi_Y = Q\right\}$



The p-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite pth moments)

- **Coupling:** $\Pi(P,Q) = \Big\{\pi_{X,Y} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \ : \ \pi_X = P \ \& \ \pi_Y = Q\Big\}$
- **Cost:** $c(x,y) = \|x-y\|^p$ for transporting x to y

The p-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite pth moments)

- **Coupling:** $\Pi(P,Q) = \left\{\pi_{X,Y} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \;:\; \pi_X = P \;\&\; \pi_Y = Q\right\}$
- **Cost:** $c(x,y) = \|x-y\|^p$ for transporting x to y

Definition (p-Wasserstein distance)

The p-Wasserstein distance: $\mathsf{W}_p(P,Q) := \left(\inf_{\pi_{X,Y} \in \Pi(P,Q)} \mathbb{E}_\pi\left[\|X-Y\|^p\right]\right)^{1/p}$

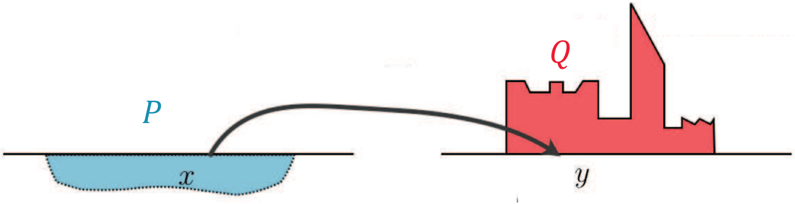
The p-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite pth moments)

- **Coupling:** $\Pi(P,Q) = \left\{\pi_{X,Y} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \ : \ \pi_X = P \ \& \ \pi_Y = Q\right\}$
- **Cost:** $c(x,y) = \|x-y\|^p$ for transporting x to y

Definition (p-Wasserstein distance)

The p-Wasserstein distance: $\mathsf{W}_p(P,Q) := \left(\inf_{\pi_{X,Y}\in\Pi(P,Q)} \mathbb{E}_\pi\left[\|X-Y\|^p\right]\right)^{1/p}$



The p-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite pth moments)

- **Coupling:** $\Pi(P,Q) = \left\{\pi_{X,Y} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \ : \ \pi_X = P \ \& \ \pi_Y = Q\right\}$
- **Cost:** $c(x,y) = \|x-y\|^p$ for transporting x to y

Definition (p-Wasserstein distance)

The p-Wasserstein distance: $\mathsf{W}_p(P,Q) := \left(\inf_{\pi_{X,Y}\in\Pi(P,Q)} \mathbb{E}_\pi\left[\|X-Y\|^p\right]\right)^{1/p}$

Comments:

The p-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite pth moments)

- **Coupling:** $\Pi(P,Q) = \left\{\pi_{X,Y} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \ : \ \pi_X = P \ \& \ \pi_Y = Q\right\}$
- **Cost:** $c(x,y) = \|x-y\|^p$ for transporting x to y

Definition (p-Wasserstein distance)

The p-Wasserstein distance: $\mathsf{W}_p(P,Q) := \left(\inf_{\pi_{X,Y} \in \Pi(P,Q)} \mathbb{E}_\pi\left[\|X-Y\|^p\right]\right)^{1/p}$

Comments:

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

The p-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite pth moments)

- **Coupling:** $\Pi(P,Q) = \left\{\pi_{X,Y} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \ : \ \pi_X = P \ \& \ \pi_Y = Q\right\}$
- **Cost:** $c(x,y) = \|x-y\|^p$ for transporting x to y

Definition (p-Wasserstein distance)

The p-Wasserstein distance: $\mathsf{W}_p(P,Q) := \left(\inf_{\pi_{X,Y} \in \Pi(P,Q)} \mathbb{E}_\pi\left[\|X-Y\|^p\right]\right)^{1/p}$

Comments:

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

The p-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$ (subscript for finite pth moments)

- **Coupling:** $\Pi(P,Q) = \left\{\pi_{X,Y} \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d) \ : \ \pi_X = P \ \& \ \pi_Y = Q\right\}$
- **Cost:** $c(x,y) = \|x-y\|^p$ for transporting x to y

Definition (p-Wasserstein distance)

The p-Wasserstein distance: $\mathsf{W}_p(P,Q) := \left(\inf_{\pi_{X,Y}\in\Pi(P,Q)} \mathbb{E}_\pi\left[\|X-Y\|^p\right]\right)^{1/p}$

Comments:

- **Robust to support mismatch:** $\mathsf{W}_p(P,Q) < \infty$, $\forall P, Q \in \mathcal{P}_p(\mathbb{R}^d)$
- **Metric:** $\left(\mathcal{P}_p(\mathbb{R}^d), \mathsf{W}_p\right)$ is metric space (metrizes weak convergence)
- **Duality:** $\mathsf{W}_1(P,Q) = \sup_{f\in\mathsf{Lip}_1(\mathbb{R}^d)} \mathbb{E}_P[f] - \mathbb{E}_Q[f] \implies$ **W-GAN** (minimax)

From Duality to Generative Adversarial Networks

Dual Representation: $$\mathsf{W}_1(P,Q) = \sup_{f\in\mathsf{Lip}_1(\mathbb{R}^d)} \mathbb{E}_P\big[f(X)\big] - \mathbb{E}_Q\big[f(Y)\big]$$

From Duality to Generative Adversarial Networks

Dual Representation: $\mathsf{W}_1(P,Q) = \sup_{f \in \mathsf{Lip}_1(\mathbb{R}^d)} \mathbb{E}_P\big[f(X)\big] - \mathbb{E}_Q\big[f(Y)\big]$

W-GANs [Arjovsky *et al*'17]:

From Duality to Generative Adversarial Networks

Dual Representation: $\mathsf{W}_1(\boldsymbol{P}, Q) = \sup_{f \in \mathsf{Lip}_1(\mathbb{R}^d)} \mathbb{E}_{\boldsymbol{P}}\big[f(\boldsymbol{X})\big] - \mathbb{E}_Q\big[f(Y)\big]$

W-GANs [Arjovsky *et al*'17]:

- P (X (real) data sample)

From Duality to Generative Adversarial Networks

Dual Representation: $\mathsf{W}_1(\boldsymbol{P}, \boldsymbol{Q}) = \sup_{f \in \mathsf{Lip}_1(\mathbb{R}^d)} \mathbb{E}_{\boldsymbol{P}}\big[f(\boldsymbol{X})\big] - \mathbb{E}_{\boldsymbol{Q}}\big[f(\boldsymbol{Y})\big]$

W-GANs [Arjovsky *et al*'17]:

- P (X (real) data sample)
- $Q = Q_\theta$ ($Y = g_\theta(Z)$ gen. sample)

From Duality to Generative Adversarial Networks

Dual Representation: $\mathsf{W}_1(\boldsymbol{P}, \boldsymbol{Q}) = \sup_{\boldsymbol{f} \in \mathsf{Lip}_1(\mathbb{R}^d)} \mathbb{E}_{\boldsymbol{P}}[\boldsymbol{f}(\boldsymbol{X})] - \mathbb{E}_{\boldsymbol{Q}}[\boldsymbol{f}(\boldsymbol{Y})]$

W-GANs [Arjovsky *et al*'17]:

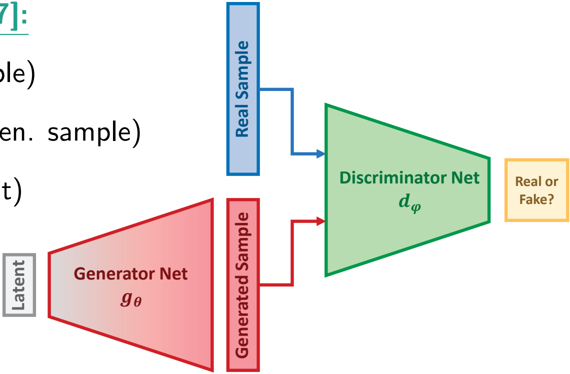
- P (X (real) data sample)
- $Q = Q_\theta$ ($Y = g_\theta(Z)$ gen. sample)
- $f = d_\varphi$ (Lip_1 constraint)

From Duality to Generative Adversarial Networks

Dual Representation: $$\mathrm{W}_1(\boldsymbol{P},\boldsymbol{Q}) = \sup_{\boldsymbol{f}\in \mathrm{Lip}_1(\mathbb{R}^d)} \mathbb{E}_{\boldsymbol{P}}[\boldsymbol{f}(\boldsymbol{X})] - \mathbb{E}_{\boldsymbol{Q}}[\boldsymbol{f}(\boldsymbol{Y})]$$

W-GANs [Arjovsky *et al*'17]:

- P (X (real) data sample)
- $Q = Q_\theta$ ($Y = g_\theta(Z)$ gen. sample)
- $f = d_\varphi$ (Lip_1 constraint)

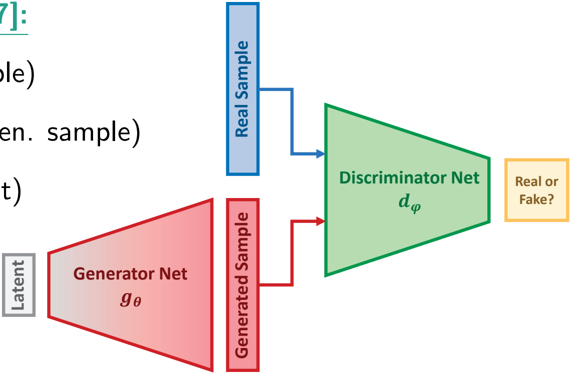


From Duality to Generative Adversarial Networks

Dual Representation: $\mathsf{W}_1(\boldsymbol{P}, \boldsymbol{Q}) = \sup_{\boldsymbol{f} \in \mathsf{Lip}_1(\mathbb{R}^d)} \mathbb{E}_{\boldsymbol{P}}[\boldsymbol{f}(\boldsymbol{X})] - \mathbb{E}_{\boldsymbol{Q}}[\boldsymbol{f}(\boldsymbol{Y})]$

W-GANs [Arjovsky *et al*'17]:

- P (X (real) data sample)
- $Q = Q_\theta$ ($Y = g_\theta(Z)$ gen. sample)
- $f = d_\varphi$ (Lip_1 constraint)



$$\Longrightarrow \quad \inf_\theta \mathsf{W}_1(P, Q_\theta) \cong \inf_\theta \sup_{\varphi:\ d_\varphi \in \mathsf{Lip}_1(\mathbb{R}^d)} \mathbb{E}[d_\varphi(X)] - \mathbb{E}[d_\varphi(g_\theta(Z))]$$

Generative Adversarial Networks

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

Implicit Generative Models: Generalization

Goal: Solve $\mathsf{OPT} := \inf_{\theta} \mathsf{W}_1 \left(P, Q_{\theta} \right)$ exactly (find $\theta^{\star}$)

Implicit Generative Models: Generalization

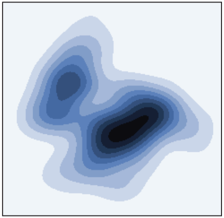
Goal: Solve $\mathsf{OPT} := \inf_\theta \mathsf{W}_1(P, Q_\theta)$ exactly (find $\theta^\star$)

Estimation: We don't have P but data

Implicit Generative Models: Generalization

Goal: Solve OPT $:= \inf_\theta \mathsf{W}_1(P, Q_\theta)$ exactly (find $\theta^\star$)

Estimation: We don't have P but data

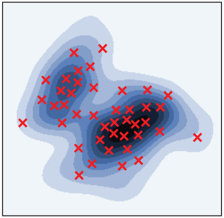


- $\{X_i\}_{i=1}^n$ are i.i.d. samples from $P \in \mathcal{P}(\mathbb{R}^d)$

Implicit Generative Models: Generalization

Goal: Solve OPT $:= \inf_{\theta} \mathsf{W}_1(P, Q_{\theta})$ exactly (find $\theta^{\star}$)

Estimation: We don't have P but data



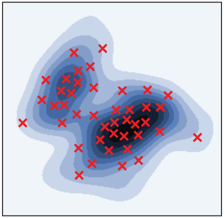
- $\{X_i\}_{i=1}^{n}$ are i.i.d. samples from $P \in \mathcal{P}(\mathbb{R}^d)$
- Empirical distribution $P_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}$

Implicit Generative Models: Generalization

Goal: Solve OPT $:= \inf_\theta \mathsf{W}_1(P, Q_\theta)$ exactly (find $\theta^\star$)

Estimation: We don't have P but data

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

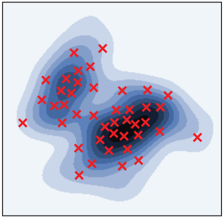


$\implies$ Inherently we work with $\mathsf{W}_1(P_n, Q_\theta)$

Implicit Generative Models: Generalization

Goal: Solve OPT $:= \inf_\theta \mathsf{W}_1(P, Q_\theta)$ exactly (find $\theta^\star$)

Estimation: We don't have P but data



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

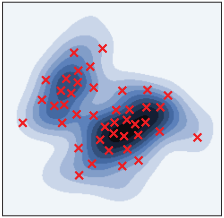
$\implies$ Inherently we work with $\mathsf{W}_1(P_n, Q_\theta)$

Optimization: Can solve $\inf_\theta \mathsf{W}_1(P_n, Q_\theta)$ approximately

Implicit Generative Models: Generalization

Goal: Solve OPT $:= \inf_\theta \mathsf{W}_1(P, Q_\theta)$ exactly (find $\theta^\star$)

Estimation: We don't have P but data



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

$\Longrightarrow$ Inherently we work with $\mathsf{W}_1(P_n, Q_\theta)$

Optimization: Can solve $\inf_\theta \mathsf{W}_1(P_n, Q_\theta)$ approximately

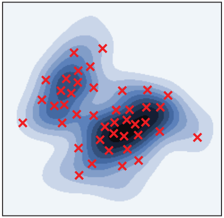
$$\text{Find } \hat{\theta}_n \text{ s.t. } \mathsf{W}_1(P_n, Q_{\hat{\theta}_n}) \le \inf_\theta \mathsf{W}_1(P_n, Q_\theta) + \epsilon$$

Implicit Generative Models: Generalization

Goal: Solve OPT $:= \inf_\theta \mathsf{W}_1(P, Q_\theta)$ exactly (find $\theta^\star$)

Estimation: We don't have P but data

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



$\Longrightarrow$ Inherently we work with $\mathsf{W}_1(P_n, Q_\theta)$

Optimization: Can solve $\inf_\theta \mathsf{W}_1(P_n, Q_\theta)$ approximately

Find $\hat{\theta}_n$ s.t. $\mathsf{W}_1(P_n, Q_{\hat{\theta}_n}) \le \inf_\theta \mathsf{W}_1(P_n, Q_\theta) + \epsilon$

Generalization: $\mathsf{W}_1(P, Q_{\hat{\theta}_n}) -$ OPT

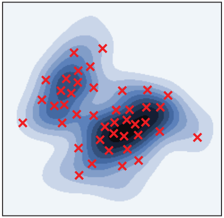
Implicit Generative Models: Generalization

Goal: Solve OPT $:= \inf_\theta \mathsf{W}_1(P, Q_\theta)$ exactly (find $\theta^\star$)

Estimation: We don't have P but data

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

$\Longrightarrow$ Inherently we work with $\mathsf{W}_1(P_n, Q_\theta)$



Optimization: Can solve $\inf_\theta \mathsf{W}_1(P_n, Q_\theta)$ approximately

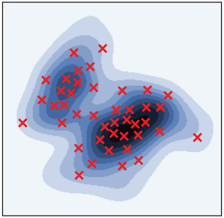
Find $\hat{\theta}_n$ s.t. $\mathsf{W}_1(P_n, Q_{\hat{\theta}_n}) \leq \inf_\theta \mathsf{W}_1(P_n, Q_\theta) + \epsilon$

Generalization: $\mathsf{W}_1(P, Q_{\hat{\theta}_n}) - \text{OPT} \leq 2\mathsf{W}_1(P_n, P) + \epsilon$

Implicit Generative Models: Generalization

Goal: Solve OPT $:= \inf_\theta \mathsf{W}_1(P, Q_\theta)$ exactly (find $\theta^\star$)

Estimation: We don't have P but data



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

$\Longrightarrow$ Inherently we work with $\mathsf{W}_1(P_n, Q_\theta)$

Optimization: Can solve $\inf_\theta \mathsf{W}_1(P_n, Q_\theta)$ approximately

$$\text{Find } \hat{\theta}_n \text{ s.t. } \mathsf{W}_1(P_n, Q_{\hat{\theta}_n}) \leq \inf_\theta \mathsf{W}_1(P_n, Q_\theta) + \epsilon$$

Generalization: $\mathsf{W}_1(P, Q_{\hat{\theta}_n}) - \text{OPT} \leq 2\mathsf{W}_1(P_n, P) + \epsilon$

$\Longrightarrow$ **Boils down to empirical approximation question under W_1**

Empirical Approximation in High Dimensions

Question: What can we say about $\mathsf{W}_1(P_n, P)$?

Empirical Approximation in High Dimensions

Question: What can we say about $\mathsf{W}_1(P_n, P)$?

Theorem (Dudley'69)

For $d \geq 3$ *and* $P \in \mathcal{P}_1(\mathbb{R}^d)$ *with* $P \ll \mathsf{Leb}(\mathbb{R}^d)$*:*

$$\mathbb{E}\big[\mathsf{W}_1(P_n, P)\big] \asymp n^{-\frac{1}{d}}$$

Empirical Approximation in High Dimensions

Question: What can we say about $\mathsf{W}_1(P_n, P)$?

Theorem (Dudley'69)

For $d \geq 3$ and $P \in \mathcal{P}_1(\mathbb{R}^d)$ with $P \ll \mathsf{Leb}(\mathbb{R}^d)$:

$$\mathbb{E}\big[\mathsf{W}_1(P_n, P)\big] \asymp n^{-\frac{1}{d}}$$

⊛ **Implication:** Too slow given dimensionality of real-world data

Empirical Approximation in High Dimensions

Question: What can we say about $\mathsf{W}_1(P_n, P)$?

Theorem (Dudley'69)

For $d \geq 3$ and $P \in \mathcal{P}_1(\mathbb{R}^d)$ with $P \ll \mathsf{Leb}(\mathbb{R}^d)$:

$$\mathbb{E}\big[\mathsf{W}_1(P_n, P)\big] \asymp n^{-\frac{1}{d}}$$

⊛ **Implication:** Too slow given dimensionality of real-world data

⊛ **Question:** How to preserve Wasserstein structure but alleviates CoD?

Smooth 1-Wasserstein Distance

Definition (Goldfeld-Greenewald'20)

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

$$\mathsf{W}_1^{(\sigma)}(P,Q) := \mathsf{W}_1(P * \mathcal{N}_\sigma, Q * \mathcal{N}_\sigma),$$

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

Smooth 1-Wasserstein Distance

Definition (Goldfeld-Greenewald'20)

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

$$\mathsf{W}_1^{(\sigma)}(P,Q) := \mathsf{W}_1(P * \mathcal{N}_\sigma, Q * \mathcal{N}_\sigma),$$

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

Interpretation: $X \sim P$, $Y \sim Q$ and $Z_1, Z_2 \sim \mathcal{N}_\sigma$

Smooth 1-Wasserstein Distance

Definition (Goldfeld-Greenewald'20)

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

$$\mathsf{W}_1^{(\sigma)}(P,Q) := \mathsf{W}_1(P * \mathcal{N}_\sigma, Q * \mathcal{N}_\sigma),$$

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

Interpretation: $X \sim P$, $Y \sim Q$ and $Z_1, Z_2 \sim \mathcal{N}_\sigma$

$$X \perp Z_1 \implies X + Z_1 \sim P * \mathcal{N}_\sigma \quad \& \quad Y \perp Z_2 \implies Y + Z_2 \sim Q * \mathcal{N}_\sigma$$

Smooth 1-Wasserstein Distance

Definition (Goldfeld-Greenewald'20)

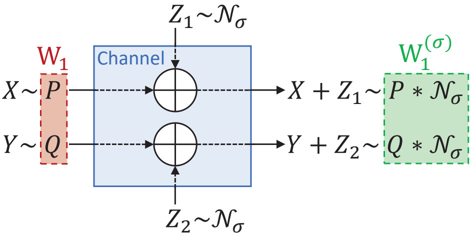
For $\sigma \geq 0$, the smooth 1-Wasserstein distance between P and Q is

$$\mathsf{W}_1^{(\sigma)}(P,Q) := \mathsf{W}_1(P * \mathcal{N}_\sigma, Q * \mathcal{N}_\sigma),$$

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

Interpretation: $X \sim P$, $Y \sim Q$ and $Z_1, Z_2 \sim \mathcal{N}_\sigma$

$$X \perp Z_1 \implies X + Z_1 \sim P * \mathcal{N}_\sigma \quad \& \quad Y \perp Z_2 \implies Y + Z_2 \sim Q * \mathcal{N}_\sigma$$



Smooth 1-Wasserstein Distance

Definition (Goldfeld-Greenewald'20)

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

$$\mathsf{W}_1^{(\sigma)}(P,Q) := \mathsf{W}_1(P * \mathcal{N}_\sigma, Q * \mathcal{N}_\sigma),$$

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

Interpretation: $X \sim P$, $Y \sim Q$ and $Z_1, Z_2 \sim \mathcal{N}_\sigma$

$$X \perp Z_1 \implies X + Z_1 \sim P * \mathcal{N}_\sigma \quad \& \quad Y \perp Z_2 \implies Y + Z_2 \sim Q * \mathcal{N}_\sigma$$

Retains KR duality: $\mathsf{W}_1^{(\sigma)}$ is W_1 but between convolved distributions:

$$\mathsf{W}_1^{(\sigma)}(P,Q) = \sup_{f \in \mathsf{Lip}_1(\mathbb{R}^d)} \mathbb{E}\big[f(X+Z)\big] - \mathbb{E}\big[f(Y+Z)\big]$$

Smooth 1-Wasserstein: Metric Structure

High Level: $\mathsf{W}_1^{(\sigma)}$ inherits the metric structure of 1-Wasserstein

Smooth 1-Wasserstein: Metric Structure

High Level: $\mathsf{W}_1^{(\sigma)}$ inherits the metric structure of 1-Wasserstein

Theorem (ZG-Greenewald'20)

$\left(\mathcal{P}_1(\mathbb{R}^d), \mathsf{W}_1^{(\sigma)}\right)$ *is metric space,* $\forall \sigma \geq 0$ *(and* $\mathsf{W}_1^{(\sigma)}$ *metrizes weak conv.).*

Smooth 1-Wasserstein: Metric Structure

High Level: $\mathsf{W}_1^{(\sigma)}$ inherits the metric structure of 1-Wasserstein

Theorem (ZG-Greenewald'20)

$\left(\mathcal{P}_1(\mathbb{R}^d), \mathsf{W}_1^{(\sigma)}\right)$ *is metric space,* $\forall \sigma \geq 0$ *(and* $\mathsf{W}_1^{(\sigma)}$ *metrizes weak conv.).*

Key Idea for Pf.: Use characteristic functions $\Phi_P(t) \triangleq \mathbb{E}_P\left[e^{itX}\right]$ and:

Smooth 1-Wasserstein: Metric Structure

High Level: $\mathsf{W}_1^{(\sigma)}$ inherits the metric structure of 1-Wasserstein

Theorem (ZG-Greenewald'20)

$\left(\mathcal{P}_1(\mathbb{R}^d), \mathsf{W}_1^{(\sigma)}\right)$ *is metric space,* $\forall \sigma \geq 0$ *(and* $\mathsf{W}_1^{(\sigma)}$ *metrizes weak conv.).*

Key Idea for Pf.: Use characteristic functions $\Phi_P(t) \triangleq \mathbb{E}_P\left[e^{itX}\right]$ and:

$$\Phi_{P*\mathcal{N}_\sigma}(t) = \Phi_P(t)\Phi_{\mathcal{N}_\sigma}(t) \text{ together with } \Phi_{\mathcal{N}_\sigma}(t) = e^{-\frac{\sigma^2\|t\|^2}{2}} \neq 0, \ \forall t.$$

Smooth 1-Wasserstein: Metric Structure

High Level: $\mathsf{W}_1^{(\sigma)}$ inherits the metric structure of 1-Wasserstein

Theorem (ZG-Greenewald'20)

$\left(\mathcal{P}_1(\mathbb{R}^d), \mathsf{W}_1^{(\sigma)}\right)$ *is metric space,* $\forall \sigma \geq 0$ *(and* $\mathsf{W}_1^{(\sigma)}$ *metrizes weak conv.).*

Key Idea for Pf.: Use characteristic functions $\Phi_P(t) \triangleq \mathbb{E}_P\left[e^{itX}\right]$ and:

$$\Phi_{P*\mathcal{N}_\sigma}(t) = \Phi_P(t)\Phi_{\mathcal{N}_\sigma}(t) \text{ together with } \Phi_{\mathcal{N}_\sigma}(t) = e^{-\frac{\sigma^2\|t\|^2}{2}} \neq 0,\ \forall t.$$

Corollary (ZG-Greenewald'20)

Let $P_n, P \in \mathcal{P}(\mathbb{R}^d)$, $n \geq 1$. *Then:* $\mathsf{W}_1^{(\sigma)}(P_n, P) \to 0$ *iff* $\mathsf{W}_1(P_n, P) \to 0$

Smooth 1-Wasserstein: Metric Structure

High Level: $\mathsf{W}_1^{(\sigma)}$ inherits the metric structure of 1-Wasserstein

Theorem (ZG-Greenewald'20)

$\left(\mathcal{P}_1(\mathbb{R}^d), \mathsf{W}_1^{(\sigma)}\right)$ *is metric space,* $\forall \sigma \geq 0$ *(and* $\mathsf{W}_1^{(\sigma)}$ *metrizes weak conv.).*

Key Idea for Pf.: Use characteristic functions $\Phi_P(t) \triangleq \mathbb{E}_P\left[e^{itX}\right]$ and:

$$\Phi_{P*\mathcal{N}_\sigma}(t) = \Phi_P(t)\Phi_{\mathcal{N}_\sigma}(t) \text{ together with } \Phi_{\mathcal{N}_\sigma}(t) = e^{-\frac{\sigma^2\|t\|^2}{2}} \neq 0,\ \forall t.$$

Corollary (ZG-Greenewald'20)

Let $P_n, P \in \mathcal{P}(\mathbb{R}^d)$, $n \geq 1$. *Then:* $\mathsf{W}_1^{(\sigma)}(P_n, P) \to 0$ *iff* $\mathsf{W}_1(P_n, P) \to 0$

⊛ $\mathsf{W}_1^{(\sigma)}$ and W_1 induce same topology

Smooth 1-Wasserstein: Function of σ

High Level: $\mathsf{W}_1^{(\sigma)}(P,Q)$ is well-behaved func. of σ (fixed $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$)

Smooth 1-Wasserstein: Function of σ

High Level: $\mathsf{W}_1^{(\sigma)}(P,Q)$ is well-behaved func. of σ (fixed $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$)

Theorem (ZG-Greenewald'20)

Fix $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$. The following hold:

Smooth 1-Wasserstein: Function of σ

High Level: $\mathsf{W}_1^{(\sigma)}(P,Q)$ is well-behaved func. of σ (fixed $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$)

Theorem (ZG-Greenewald'20)

Fix $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$. The following hold:

1. $\mathsf{W}_1^{(\sigma)}(P,Q)$ *is continuous and mono. non-increasing in* $\sigma \in [0,+\infty)$

Smooth 1-Wasserstein: Function of σ

High Level: $\mathsf{W}_1^{(\sigma)}(P,Q)$ is well-behaved func. of σ (fixed $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$)

Theorem (ZG-Greenewald'20)

Fix $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$. The following hold:

1. $\mathsf{W}_1^{(\sigma)}(P,Q)$ *is continuous and mono. non-increasing in* $\sigma \in [0,+\infty)$
2. $\lim_{\sigma\to 0} \mathsf{W}_1^{(\sigma)}(P,Q) = \mathsf{W}_1(P,Q)$

Smooth 1-Wasserstein: Function of σ

High Level: $\mathsf{W}_1^{(\sigma)}(P,Q)$ is well-behaved func. of σ (fixed $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$)

Theorem (ZG-Greenewald'20)

Fix $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$. The following hold:

1. $\mathsf{W}_1^{(\sigma)}(P,Q)$ *is continuous and mono. non-increasing in* $\sigma \in [0,+\infty)$
2. $\lim_{\sigma\to 0} \mathsf{W}_1^{(\sigma)}(P,Q) = \mathsf{W}_1(P,Q)$
3. $\lim_{\sigma\to\infty} \mathsf{W}_1^{(\sigma)}(P,Q) \neq 0$, *for some* $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$

Smooth 1-Wasserstein: Function of σ

High Level: $\mathsf{W}_1^{(\sigma)}(P,Q)$ is well-behaved func. of σ (fixed $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$)

Theorem (ZG-Greenewald'20)

Fix $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$. The following hold:

1. $\mathsf{W}_1^{(\sigma)}(P,Q)$ *is continuous and mono. non-increasing in* $\sigma \in [0,+\infty)$
2. $\lim_{\sigma\to 0} \mathsf{W}_1^{(\sigma)}(P,Q) = \mathsf{W}_1(P,Q)$
3. $\lim_{\sigma\to\infty} \mathsf{W}_1^{(\sigma)}(P,Q) \neq 0$, *for some* $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$

Pf. Items 1-2: Use dual form to derive stability lemma:

Smooth 1-Wasserstein: Function of σ

High Level: $\mathsf{W}_1^{(\sigma)}(P,Q)$ is well-behaved func. of σ (fixed $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$)

Theorem (ZG-Greenewald'20)

Fix $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$. The following hold:

1. $\mathsf{W}_1^{(\sigma)}(P,Q)$ *is continuous and mono. non-increasing in* $\sigma \in [0,+\infty)$
2. $\lim_{\sigma\to 0} \mathsf{W}_1^{(\sigma)}(P,Q) = \mathsf{W}_1(P,Q)$
3. $\lim_{\sigma\to\infty} \mathsf{W}_1^{(\sigma)}(P,Q) \neq 0$, *for some* $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$

Pf. Items 1-2: Use dual form to derive stability lemma:

Lemma (ZG-Greenewald'20)

For $\sigma_1 < \sigma_2:\ \mathsf{W}_1^{(\sigma_2)}(P,Q) \le \mathsf{W}_1^{(\sigma_1)}(P,Q) \le \mathsf{W}_1^{(\sigma_2)}(P,Q) + 2\sqrt{d(\sigma_2^2 - \sigma_1^2)}$

Smooth 1-Wasserstein: Function of σ

High Level: $\mathsf{W}_1^{(\sigma)}(P,Q)$ is well-behaved func. of σ (fixed $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$)

Theorem (ZG-Greenewald'20)

Fix $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$. The following hold:

1. $\mathsf{W}_1^{(\sigma)}(P,Q)$ *is continuous and mono. non-increasing in* $\sigma \in [0,+\infty)$
2. $\lim_{\sigma\to 0} \mathsf{W}_1^{(\sigma)}(P,Q) = \mathsf{W}_1(P,Q)$
3. $\lim_{\sigma\to\infty} \mathsf{W}_1^{(\sigma)}(P,Q) \neq 0$, *for some* $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$

Pf. Items 1-2: Use dual form to derive stability lemma:

Lemma (ZG-Greenewald'20)

For $\sigma_1 < \sigma_2: \; \mathsf{W}_1^{(\sigma_2)}(P,Q) \leq \mathsf{W}_1^{(\sigma_1)}(P,Q) \leq \mathsf{W}_1^{(\sigma_2)}(P,Q) + 2\sqrt{d(\sigma_2^2 - \sigma_1^2)}$

Pf. Item 3: $\quad \mathsf{W}_1^{(\sigma)}(\delta_x,\delta_y) = \mathsf{W}_1\big(\mathcal{N}(x,\sigma^2\mathrm{I}_d),\mathcal{N}(y,\sigma^2\mathrm{I}_d)\big) = \|x-y\|$

Smooth 1-Wasserstein: Statistical Efficiency

High Level: Alleviate curse of dimensionality & get limit distribution

Smooth 1-Wasserstein: Statistical Efficiency

High Level: Alleviate curse of dimensionality & get limit distribution

Theorem (ZG-Greenewald-Polyanskiy-Weed'20)

For any $d \geq 1$, $\sigma > 0$ *and sub-Gaussian* P*:* $\quad \mathbb{E}\left[\mathsf{W}_1^{(\sigma)}(P_n, P)\right] \lesssim n^{-\frac{1}{2}}$

Smooth 1-Wasserstein: Statistical Efficiency

High Level: Alleviate curse of dimensionality & get limit distribution

Theorem (ZG-Greenewald-Polyanskiy-Weed'20)

For any $d \geq 1$, $\sigma > 0$ *and sub-Gaussian* P*:* $\mathbb{E}\Big[\mathsf{W}_1^{(\sigma)}(P_n, P)\Big] \lesssim n^{-\frac{1}{2}}$

Theorem (ZG-Kato'20)

Under same assumptions: $\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P) \xrightarrow{d} \sup_{f \in \mathsf{Lip}_1(\mathbb{R}^d)} G_P^{(\sigma)}(f)$,
where $G_P^{(\sigma)}$ *is a tight centered Gaussian process in* $\ell^\infty(\mathsf{Lip}_1(\mathbb{R}^d))$ *with*
$\mathbb{E}\left[G_P^{(\sigma)}(f)G_P^{(\sigma)}(g)\right] = \mathsf{Cov}_P(f * \varphi_\sigma, g * \varphi_\sigma), \quad \forall f, g, \in \mathsf{Lip}_1(\mathbb{R}^d)$

Smooth 1-Wasserstein: Statistical Efficiency

High Level: Alleviate curse of dimensionality & get limit distribution

Theorem (ZG-Greenewald-Polyanskiy-Weed'20)

For any $d \geq 1$, $\sigma > 0$ *and sub-Gaussian* P: $\quad \mathbb{E}\Big[\mathsf{W}_1^{(\sigma)}(P_n, P)\Big] \lesssim n^{-\frac{1}{2}}$

Theorem (ZG-Kato'20)

Under same assumptions: $\quad \sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P) \xrightarrow{d} \sup_{f \in \mathsf{Lip}_1(\mathbb{R}^d)} G_P^{(\sigma)}(f)$,
where $G_P^{(\sigma)}$ *is a tight centered Gaussian process in* $\ell^\infty(\mathsf{Lip}_1(\mathbb{R}^d))$ *with*
$\mathbb{E}\left[G_P^{(\sigma)}(f) G_P^{(\sigma)}(g)\right] = \mathsf{Cov}_P(f * \varphi_\sigma, g * \varphi_\sigma), \quad \forall f, g, \in \mathsf{Lip}_1(\mathbb{R}^d)$

Comments:

- Limit distribution shows $n^{-1/2}$ rate is sharp

Smooth 1-Wasserstein: Statistical Efficiency

High Level: Alleviate curse of dimensionality & get limit distribution

Theorem (ZG-Greenewald-Polyanskiy-Weed'20)

For any $d \geq 1$, $\sigma > 0$ and sub-Gaussian P: $\mathbb{E}\Big[\mathsf{W}_1^{(\sigma)}(P_n, P)\Big] \lesssim n^{-\frac{1}{2}}$

Theorem (ZG-Kato'20)

Under same assumptions: $\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P) \stackrel{d}{\to} \sup_{f \in \mathsf{Lip}_1(\mathbb{R}^d)} G_P^{(\sigma)}(f)$,
where $G_P^{(\sigma)}$ is a tight centered Gaussian process in $\ell^\infty(\mathsf{Lip}_1(\mathbb{R}^d))$ with
$\mathbb{E}\left[G_P^{(\sigma)}(f) G_P^{(\sigma)}(g)\right] = \mathsf{Cov}_P(f * \varphi_\sigma, g * \varphi_\sigma), \quad \forall f, g, \in \mathsf{Lip}_1(\mathbb{R}^d)$

Comments:

- Limit distribution shows $n^{-1/2}$ rate is sharp
- W_1 limit dist. known only for $d = 1$ ($\mathsf{W}_1(P_n, P) = \|F_n - F\|_{L^1(\mathbb{R})}$)

Smooth 1-Wasserstein: Statistical Efficiency

High Level: Alleviate curse of dimensionality & get limit distribution

Theorem (ZG-Greenewald-Polyanskiy-Weed'20)

For any $d \geq 1$, $\sigma > 0$ and sub-Gaussian P: $\mathbb{E}\Big[\mathsf{W}_1^{(\sigma)}(P_n, P)\Big] \lesssim n^{-\frac{1}{2}}$

Theorem (ZG-Kato'20)

Under same assumptions: $\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P) \xrightarrow{d} \sup_{f \in \mathsf{Lip}_1(\mathbb{R}^d)} G_P^{(\sigma)}(f)$,
where $G_P^{(\sigma)}$ is a tight centered Gaussian process in $\ell^\infty(\mathsf{Lip}_1(\mathbb{R}^d))$ with
$\mathbb{E}\left[G_P^{(\sigma)}(f) G_P^{(\sigma)}(g)\right] = \mathsf{Cov}_P(f * \varphi_\sigma, g * \varphi_\sigma), \quad \forall f, g, \in \mathsf{Lip}_1(\mathbb{R}^d)$

Comments:

- Limit distribution shows $n^{-1/2}$ rate is sharp
- W_1 limit dist. known only for $d = 1$ ($\mathsf{W}_1(P_n, P) = \|F_n - F\|_{L^1(\mathbb{R})}$)
- $\mathsf{W}_1^{(\sigma)}$ has a stable limit (characterized) in all d!

Limit Distribution Proof Roadmap

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

Limit Distribution Proof Roadmap

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

$$\mathsf{W}_1^{(\sigma)}(P,Q) = \sup_{f \in \mathsf{Lip}_1(\mathbb{R}^d)} \mathbb{E}_P[f * \varphi_\sigma] - \mathbb{E}_Q[f * \varphi_\sigma] = \sup_{g \in \mathcal{F}_\sigma} \mathbb{E}_P[g] - \mathbb{E}_Q[g]$$

Limit Distribution Proof Roadmap

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

$$\mathsf{W}_1^{(\sigma)}(P,Q) = \sup_{f \in \mathsf{Lip}_1(\mathbb{R}^d)} \mathbb{E}_P[f * \varphi_\sigma] - \mathbb{E}_Q[f * \varphi_\sigma] = \sup_{g \in \mathcal{F}_\sigma} \mathbb{E}_P[g] - \mathbb{E}_Q[g]$$

$\Longrightarrow$ **Emp.** $\mathbf{W_1^{(\sigma)}}$**:**

$$\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n,P) = \sqrt{n} \sup_{g \in \mathcal{F}_\sigma} \underbrace{\frac{1}{n}\sum_{i=1}^n g(X_i) - \mathbb{E}_P[g]}_{=:(P_n - P)g}$$

Limit Distribution Proof Roadmap

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

$$\mathsf{W}_1^{(\sigma)}(P,Q) = \sup_{f \in \mathsf{Lip}_1(\mathbb{R}^d)} \mathbb{E}_P[f * \varphi_\sigma] - \mathbb{E}_Q[f * \varphi_\sigma] = \sup_{g \in \mathcal{F}_\sigma} \mathbb{E}_P[g] - \mathbb{E}_Q[g]$$

$\Longrightarrow$ **Emp. $\mathbf{W}_1^{(\sigma)}$:**

$$\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P) = \sqrt{n} \sup_{g \in \mathcal{F}_\sigma} \underbrace{\frac{1}{n}\sum_{i=1}^n g(X_i) - \mathbb{E}_P[g]}_{=:(P_n - P)g}$$

⊛ **Objective:** $\mathbb{G}_n = \left(\sqrt{n}(P_n - P)g\right)_{g \in \mathcal{F}_\sigma} \xrightarrow{w} \mathsf{GP}$

Limit Distribution Proof Roadmap

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

$$\mathsf{W}_1^{(\sigma)}(P,Q) = \sup_{f\in\mathsf{Lip}_1(\mathbb{R}^d)} \mathbb{E}_P[f*\varphi_\sigma] - \mathbb{E}_Q[f*\varphi_\sigma] = \sup_{g\in\mathcal{F}_\sigma} \mathbb{E}_P[g] - \mathbb{E}_Q[g]$$

$\implies$ **Emp. $\mathsf{W}_1^{(\sigma)}$:**

$$\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n,P) = \sqrt{n}\sup_{g\in\mathcal{F}_\sigma} \underbrace{\frac{1}{n}\sum_{i=1}^n g(X_i) - \mathbb{E}_P[g]}_{=:(P_n-P)g}$$

⊛ **Objective:** $\mathbb{G}_n = \left(\sqrt{n}(P_n - P)g\right)_{g\in\mathcal{F}_\sigma} \xrightarrow{w}$ GP $\iff$ $\mathcal{F}_\sigma$ is $P-$Donsker

Limit Distribution Proof Roadmap

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

$$\mathsf{W}_1^{(\sigma)}(P,Q) = \sup_{f \in \mathsf{Lip}_1(\mathbb{R}^d)} \mathbb{E}_P[f * \varphi_\sigma] - \mathbb{E}_Q[f * \varphi_\sigma] = \sup_{g \in \mathcal{F}_\sigma} \mathbb{E}_P[g] - \mathbb{E}_Q[g]$$

$\Longrightarrow$ **Emp. $\mathsf{W}_1^{(\sigma)}$:**

$$\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P) = \sqrt{n} \sup_{g \in \mathcal{F}_\sigma} \underbrace{\frac{1}{n}\sum_{i=1}^n g(X_i) - \mathbb{E}_P[g]}_{=:(P_n - P)g}$$

⊛ **Objective:** $\mathbb{G}_n = \left(\sqrt{n}(P_n - P)g\right)_{g \in \mathcal{F}_\sigma} \xrightarrow{w} \mathsf{GP} \iff \mathcal{F}_\sigma$ is $P-$Donsker

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

Limit Distribution Proof Roadmap

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

$$\mathsf{W}_1^{(\sigma)}(P,Q) = \sup_{f \in \mathsf{Lip}_1(\mathbb{R}^d)} \mathbb{E}_P[f * \varphi_\sigma] - \mathbb{E}_Q[f * \varphi_\sigma] = \sup_{g \in \mathcal{F}_\sigma} \mathbb{E}_P[g] - \mathbb{E}_Q[g]$$

$\Longrightarrow$ **Emp. $\mathbf{W_1^{(\sigma)}}$:**

$$\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P) = \sqrt{n} \sup_{g \in \mathcal{F}_\sigma} \underbrace{\frac{1}{n}\sum_{i=1}^n g(X_i) - \mathbb{E}_P[g]}_{=:(P_n - P)g}$$

⊛ **Objective:** $\mathbb{G}_n = \left(\sqrt{n}(P_n - P)g\right)_{g \in \mathcal{F}_\sigma} \xrightarrow{w} \mathsf{GP} \iff \mathcal{F}_\sigma$ is $P-$Donsker

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

$\Longrightarrow$ **$\boldsymbol{\mathcal{F}_\sigma}$ is $\boldsymbol{P}$-Donsker:**

$$\mathbb{G}_n \xrightarrow{w} G_P \text{ (Gaussian process)}$$

Limit Distribution Proof Roadmap

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

$$\mathsf{W}_1^{(\sigma)}(P,Q) = \sup_{f \in \mathsf{Lip}_1(\mathbb{R}^d)} \mathbb{E}_P[f * \varphi_\sigma] - \mathbb{E}_Q[f * \varphi_\sigma] = \sup_{g \in \mathcal{F}_\sigma} \mathbb{E}_P[g] - \mathbb{E}_Q[g]$$

$\Longrightarrow$ **Emp. $\mathbf{W}_1^{(\sigma)}$:**

$$\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n,P) = \sqrt{n} \sup_{g \in \mathcal{F}_\sigma} \underbrace{\frac{1}{n}\sum_{i=1}^n g(X_i) - \mathbb{E}_P[g]}_{=:(P_n - P)g}$$

⊛ **Objective:** $\mathbb{G}_n = \left(\sqrt{n}(P_n - P)g\right)_{g \in \mathcal{F}_\sigma} \xrightarrow{w} \mathsf{GP} \iff \mathcal{F}_\sigma$ is $P-$Donsker

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

$\Longrightarrow$ **$\boldsymbol{\mathcal{F}_\sigma}$ is $\boldsymbol{P}$-Donsker:**

$$\mathbb{G}_n \xrightarrow{w} G_P \text{ (Gaussian process)}$$

3. **Continuous Mapping Theorem:**

$$\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n,P) \xrightarrow{d} \sup_{g \in \mathcal{F}_\sigma} G_P(g)$$

Smooth p-Wasserstein Distances

Definition (Smooth p-Wasserstein distance)

For $\sigma \geq 0$ and $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$, define $\mathsf{W}_p^{(\sigma)}(P, Q) := \mathsf{W}_p(P * \mathcal{N}_\sigma, Q * \mathcal{N}_\sigma)$.

Smooth p-Wasserstein Distances

Definition (Smooth p-Wasserstein distance)

For $\sigma \geq 0$ and $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$, define $\mathsf{W}_p^{(\sigma)}(P,Q) := \mathsf{W}_p(P * \mathcal{N}_\sigma, Q * \mathcal{N}_\sigma)$.

Structural Properties: Everything goes through

Smooth p-Wasserstein Distances

Definition (Smooth p-Wasserstein distance)

For $\sigma \geq 0$ and $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$, define $\mathsf{W}_p^{(\sigma)}(P,Q) := \mathsf{W}_p(P * \mathcal{N}_\sigma, Q * \mathcal{N}_\sigma)$.

Structural Properties: Everything goes through

1. **Metric:** $(\mathcal{P}_p(\mathbb{R}^d), \mathsf{W}_p)$ metric space & topologically equivalent to W_p

Smooth p-Wasserstein Distances

Definition (Smooth p-Wasserstein distance)

For $\sigma \geq 0$ and $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$, define $\mathsf{W}_p^{(\sigma)}(P,Q) := \mathsf{W}_p(P * \mathcal{N}_\sigma, Q * \mathcal{N}_\sigma)$.

Structural Properties: Everything goes through

1. **Metric:** $(\mathcal{P}_p(\mathbb{R}^d), \mathsf{W}_p)$ metric space **&** topologically equivalent to W_p
2. **Stability:** $\left|\mathsf{W}_p^{(\sigma)}(P,Q) - \mathsf{W}_p(P,Q)\right| \leq 2\sigma\sqrt{d + 2(p-1)}$

Smooth p-Wasserstein Distances

Definition (Smooth p-Wasserstein distance)

For $\sigma \geq 0$ and $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$, define $\mathsf{W}_p^{(\sigma)}(P,Q) := \mathsf{W}_p(P * \mathcal{N}_\sigma, Q * \mathcal{N}_\sigma)$.

Structural Properties: Everything goes through

1. **Metric:** $(\mathcal{P}_p(\mathbb{R}^d), \mathsf{W}_p)$ metric space **&** topologically equivalent to W_p
2. **Stability:** $\left|\mathsf{W}_p^{(\sigma)}(P,Q) - \mathsf{W}_p(P,Q)\right| \leq 2\sigma\sqrt{d + 2(p-1)}$
3. **Regularity:** Cont. **&** monotone func. of σ + convergence of plans

Smooth p-Wasserstein Distances

Definition (Smooth p-Wasserstein distance)

For $\sigma \geq 0$ and $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$, define $\mathsf{W}_p^{(\sigma)}(P,Q) := \mathsf{W}_p(P * \mathcal{N}_\sigma, Q * \mathcal{N}_\sigma)$.

Structural Properties: Everything goes through

1. **Metric:** $(\mathcal{P}_p(\mathbb{R}^d), \mathsf{W}_p)$ metric space & topologically equivalent to W_p
2. **Stability:** $\left|\mathsf{W}_p^{(\sigma)}(P,Q) - \mathsf{W}_p(P,Q)\right| \leq 2\sigma\sqrt{d + 2(p-1)}$
3. **Regularity:** Cont. & monotone func. of σ + convergence of plans

⊛ **Question:** How about fast empirical convergence?

Smooth p-Wasserstein: Empirical Convergence

Dimension-Free Rate:

Smooth p-Wasserstein: Empirical Convergence

Dimension-Free Rate:

Theorem (Nietert-ZG-Kato'21)

For any $d \geq 1$, $\sigma > 0$ *and sub-Gaussian* P*:* $\mathbb{E}\left[\mathsf{W}_p^{(\sigma)}(P_n, P)\right] \lesssim n^{-\frac{1}{2p}}$

Smooth p-Wasserstein: Empirical Convergence

Dimension-Free Rate:

Theorem (Nietert-ZG-Kato'21)

For any $d \geq 1$, $\sigma > 0$ and sub-Gaussian P: $\mathbb{E}\Big[\mathsf{W}_p^{(\sigma)}(P_n, P)\Big] \lesssim n^{-\frac{1}{2p}}$

Pf Idea: Elementary arguments

1. [Prop. 7.10, Villani'03]: For $p \geq 0$: $\mathsf{W}_p^p(P,Q) \leq 2^p \int \|x\|^p \, \mathsf{d}(P-Q)(x)$

Smooth p-Wasserstein: Empirical Convergence

Dimension-Free Rate:

Theorem (Nietert-ZG-Kato'21)

For any $d \geq 1$, $\sigma > 0$ and sub-Gaussian P: $\mathbb{E}\left[\mathsf{W}_p^{(\sigma)}(P_n, P)\right] \lesssim n^{-\frac{1}{2p}}$

Pf Idea: Elementary arguments

1. [Prop. 7.10, Villani'03]: For $p \geq 0$: $\mathsf{W}_p^p(P,Q) \leq 2^p \int \|x\|^p \, \mathsf{d}(P-Q)(x)$
   - Proven via maximal (TV) coupling

Smooth p-Wasserstein: Empirical Convergence

Dimension-Free Rate:

Theorem (Nietert-ZG-Kato'21)

For any $d \geq 1$, $\sigma > 0$ and sub-Gaussian P: $\mathbb{E}\left[\mathsf{W}_p^{(\sigma)}(P_n, P)\right] \lesssim n^{-\frac{1}{2p}}$

Pf Idea: Elementary arguments

1. [Prop. 7.10, Villani'03]: For $p \geq 0$: $\mathsf{W}_p^p(P,Q) \leq 2^p \int \|x\|^p \, \mathsf{d}(P-Q)(x)$
   - Proven via maximal (TV) coupling
2. **Emp. RHS:** $\mathbb{E}\left[\int \|x\|^p \mathsf{d}(P_n - P)(x)\right] \leq \frac{1}{\sqrt{n}} \int \|x\|^p \sqrt{\mathsf{Var}_P\big(\varphi_\sigma(x - X)\big)} \, \mathsf{d}x$

Smooth p-Wasserstein: Empirical Convergence

Dimension-Free Rate:

Theorem (Nietert-ZG-Kato'21)

For any $d \geq 1$, $\sigma > 0$ and sub-Gaussian P: $\mathbb{E}\left[\mathsf{W}_p^{(\sigma)}(P_n, P)\right] \lesssim n^{-\frac{1}{2p}}$

Pf Idea: Elementary arguments

1. [Prop. 7.10, Villani'03]: For $p \geq 0$: $\mathsf{W}_p^p(P,Q) \leq 2^p \int \|x\|^p \, \mathsf{d}(P-Q)(x)$
   - Proven via maximal (TV) coupling
2. **Emp. RHS:** $\mathbb{E}\left[\int \|x\|^p \mathsf{d}(P_n - P)(x)\right] \leq \frac{1}{\sqrt{n}} \int \|x\|^p \sqrt{\mathsf{Var}_P\big(\varphi_\sigma(x - X)\big)}\, \mathsf{d}x$

⊛ **Issue:** Rate is dimension-free but suboptimal in p

Smooth p-Wasserstein: Empirical Convergence

Dimension-Free Rate:

Theorem (Nietert-ZG-Kato'21)

For any $d \geq 1$, $\sigma > 0$ and sub-Gaussian P: $\mathbb{E}\left[\mathsf{W}_p^{(\sigma)}(P_n, P)\right] \lesssim n^{-\frac{1}{2p}}$

Pf Idea: Elementary arguments

1. [Prop. 7.10, Villani'03]: For $p \geq 0$: $\mathsf{W}_p^p(P, Q) \leq 2^p \int \|x\|^p \, \mathsf{d}(P - Q)(x)$
   - Proven via maximal (TV) coupling
2. **Emp. RHS:** $\mathbb{E}\left[\int \|x\|^p \mathsf{d}(P_n - P)(x)\right] \leq \frac{1}{\sqrt{n}} \int \|x\|^p \sqrt{\mathsf{Var}_P\big(\varphi_\sigma(x - X)\big)} \, \mathsf{d}x$

⊛ **Issue:** Rate is dimension-free but suboptimal in p

⊛ **Question:** Can we get the right dependence?

Smooth p-Wasserstein vs Sobolev IPM: Definitions

Recall: $\mathsf{W}_1^{(\sigma)}(P_n, P) = \sup_{g \in \mathcal{F}_\sigma} \mathbb{E}_{P_n}[g] - \mathbb{E}_P[g]$ analysis gives $n^{-1/2}$ rate

Smooth p-Wasserstein vs Sobolev IPM: Definitions

<u>**Recall:**</u> $\mathsf{W}_1^{(\sigma)}(P_n, P) = \sup_{g \in \mathcal{F}_\sigma} \mathbb{E}_{P_n}[g] - \mathbb{E}_P[g]$ analysis gives $n^{-1/2}$ rate

⊛ **Issue:** W_p is not an IPM

Smooth p-Wasserstein vs Sobolev IPM: Definitions

Recall: $\mathsf{W}_1^{(\sigma)}(P_n, P) = \sup_{g \in \mathcal{F}_\sigma} \mathbb{E}_{P_n}[g] - \mathbb{E}_P[g]$ analysis gives $n^{-1/2}$ rate

⊛ **Issue:** W_p is not an IPM

⊛ **Strategy:** Upper bound W_p by an IPM

Smooth p-Wasserstein vs Sobolev IPM: Definitions

Recall: $\mathsf{W}_1^{(\sigma)}(P_n, P) = \sup_{g \in \mathcal{F}_\sigma} \mathbb{E}_{P_n}[g] - \mathbb{E}_P[g]$ analysis gives $n^{-1/2}$ rate

⊛ **Issue:** W_p is not an IPM

⊛ **Strategy:** Upper bound W_p by an IPM

Sobolev IPM: Defined wrt unit ball in Sobolev space

Smooth p-Wasserstein vs Sobolev IPM: Definitions

Recall: $\mathsf{W}_1^{(\sigma)}(P_n, P) = \sup_{g \in \mathcal{F}_\sigma} \mathbb{E}_{P_n}[g] - \mathbb{E}_P[g]$ analysis gives $n^{-1/2}$ rate

⊛ **Issue:** W_p is not an IPM

⊛ **Strategy:** Upper bound W_p by an IPM

Sobolev IPM: Defined wrt unit ball in Sobolev space

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

Smooth p-Wasserstein vs Sobolev IPM: Definitions

Recall: $\mathsf{W}_1^{(\sigma)}(P_n, P) = \sup_{g \in \mathcal{F}_\sigma} \mathbb{E}_{P_n}[g] - \mathbb{E}_P[g]$ analysis gives $n^{-1/2}$ rate

⊛ **Issue:** W_p is not an IPM

⊛ **Strategy:** Upper bound W_p by an IPM

Sobolev IPM: Defined wrt unit ball in Sobolev space

- **Sobolev seminorm:** $\|f\|_{\dot{H}^{1,p}(\gamma)} := \|\nabla f\|_{L^p(\gamma;\mathbb{R}^d)} = \left(\int_{\mathbb{R}^d} \|\nabla f\|^p \, \mathrm{d}\gamma\right)^{1/p}$
- **Unit ball:** $\mathcal{F}_q(\gamma) := \left\{f \in C_0^\infty(\mathbb{R}^d) : \|f\|_{\dot{H}^{1,q}(\gamma)} \le 1\right\}$, with $\frac{1}{p} + \frac{1}{q} = 1$

Smooth p-Wasserstein vs Sobolev IPM: Definitions

Recall: $\mathsf{W}_1^{(\sigma)}(P_n, P) = \sup_{g \in \mathcal{F}_\sigma} \mathbb{E}_{P_n}[g] - \mathbb{E}_P[g]$ analysis gives $n^{-1/2}$ rate

⊛ **Issue:** W_p is not an IPM

⊛ **Strategy:** Upper bound W_p by an IPM

Sobolev IPM: Defined wrt unit ball in Sobolev space

- **Sobolev seminorm:** $\|f\|_{\dot{H}^{1,p}(\gamma)} := \|\nabla f\|_{L^p(\gamma;\mathbb{R}^d)} = \left(\int_{\mathbb{R}^d} \|\nabla f\|^p \, \mathsf{d}\gamma\right)^{1/p}$
- **Unit ball:** $\mathcal{F}_q(\gamma) := \left\{f \in C_0^\infty(\mathbb{R}^d) : \|f\|_{\dot{H}^{1,q}(\gamma)} \le 1\right\}$, with $\frac{1}{p} + \frac{1}{q} = 1$

Definition (Sobolev IPM)

The pth order Sobolev IPM is: $\mathsf{d}_{\gamma,p}(P,Q) := \sup_{f \in \mathcal{F}_q(\gamma)} \mathbb{E}_P[f] - \mathbb{E}_Q[f]$.

Smooth p-Wasserstein vs Sobolev IPM: Definitions

Recall: $\mathsf{W}_1^{(\sigma)}(P_n, P) = \sup_{g \in \mathcal{F}_\sigma} \mathbb{E}_{P_n}[g] - \mathbb{E}_P[g]$ analysis gives $n^{-1/2}$ rate

⊛ **Issue:** W_p is not an IPM

⊛ **Strategy:** Upper bound W_p by an IPM

Sobolev IPM: Defined wrt unit ball in Sobolev space

- **Sobolev seminorm:** $\|f\|_{\dot{H}^{1,p}(\gamma)} := \|\nabla f\|_{L^p(\gamma;\mathbb{R}^d)} = \left(\int_{\mathbb{R}^d} \|\nabla f\|^p \, \mathsf{d}\gamma\right)^{1/p}$
- **Unit ball:** $\mathcal{F}_q(\gamma) := \{f \in C_0^\infty(\mathbb{R}^d) : \|f\|_{\dot{H}^{1,q}(\gamma)} \leq 1\}$, with $\frac{1}{p} + \frac{1}{q} = 1$

Definition (Sobolev IPM)

The pth order Sobolev IPM is: $\mathsf{d}_{\gamma,p}(P, Q) := \sup_{f \in \mathcal{F}_q(\gamma)} \mathbb{E}_P[f] - \mathbb{E}_Q[f]$.

⊛ **Question:** How do Wasserstein distances and Sobolev IPMs compare?

Smooth p-Wasserstein vs Sobolev IPM: Comparison

Theorem (Dolbeault-Nazaret-Savaré'09)

Let $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$ *satisfy* $P, Q \ll \gamma$ *with* $\frac{\mathsf{d}P}{\mathsf{d}\gamma} \geq c > 0$. *Then*

$$\mathsf{W}_p(P, Q) \leq c^{-1/q} p \, \mathsf{d}_{\gamma,p}(P, Q).$$

Smooth p-Wasserstein vs Sobolev IPM: Comparison

Theorem (Dolbeault-Nazaret-Savaré'09)

Let $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$ *satisfy* $P, Q \ll \gamma$ *with* $\frac{\mathsf{d}P}{\mathsf{d}\gamma} \geq c > 0$. *Then*

$$\mathsf{W}_p(P,Q) \leq c^{-1/q} p \, \mathsf{d}_{\gamma,p}(P,Q).$$

Comments:

Smooth p-Wasserstein vs Sobolev IPM: Comparison

Theorem (Dolbeault-Nazaret-Savaré'09)

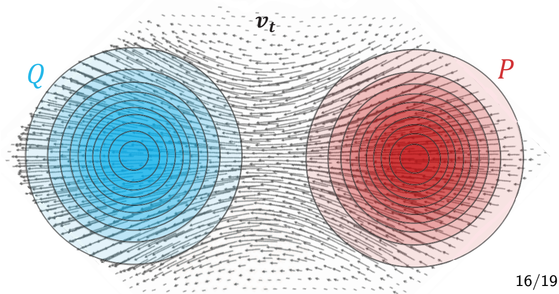
Let $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$ *satisfy* $P, Q \ll \gamma$ *with* $\frac{\mathsf{d}P}{\mathsf{d}\gamma} \geq c > 0$. *Then*

$$\mathsf{W}_p(P,Q) \leq c^{-1/q} p \, \mathsf{d}_{\gamma,p}(P,Q).$$

Comments:

1. Proof via Benamou-Brenier dynamical formulation

$$\mathsf{W}_p(P,Q) = \inf \left\{ \int_0^1 \|v_t\|_{L^p(\mu_t)} \, \mathsf{d}t : \mu_0 = P, \mu_1 = Q, \, \partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0 \right\}$$



Smooth p-Wasserstein vs Sobolev IPM: Comparison

Theorem (Dolbeault-Nazaret-Savaré'09)

Let $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$ *satisfy* $P, Q \ll \gamma$ *with* $\frac{\mathsf{d}P}{\mathsf{d}\gamma} \geq c > 0$. *Then*

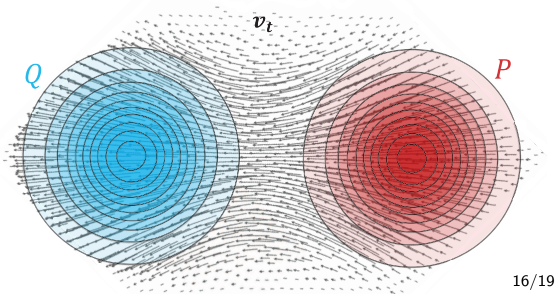
$$\mathsf{W}_p(P, Q) \leq c^{-1/q} p \, \mathsf{d}_{\gamma,p}(P, Q).$$

Comments:

1. Proof via Benamou-Brenier dynamical formulation

$$\mathsf{W}_p(P, Q) = \inf \left\{ \int_0^1 \|v_t\|_{L^p(\mu_t)} \, \mathsf{d}t : \mu_0 = P, \mu_1 = Q, \, \partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0 \right\}$$

- Take $\mu_t = (1-t)P + tQ$



Smooth p-Wasserstein vs Sobolev IPM: Comparison

Theorem (Dolbeault-Nazaret-Savaré'09)

Let $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$ *satisfy* $P, Q \ll \gamma$ *with* $\frac{\mathsf{d}P}{\mathsf{d}\gamma} \geq c > 0$. *Then*

$$\mathsf{W}_p(P,Q) \leq c^{-1/q} p\, \mathsf{d}_{\gamma,p}(P,Q).$$

Comments:

1. Proof via Benamou-Brenier dynamical formulation

$$\mathsf{W}_p(P,Q) = \inf\left\{ \int_0^1 \|v_t\|_{L^p(\mu_t)}\, \mathsf{d}t : \mu_0 = P, \mu_1 = Q,\ \partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0 \right\}$$

- Take $\mu_t = (1-t)P + tQ$

2. Lower bound automatically holds after Gaussian smoothing w/ $\gamma = \mathcal{N}_\sigma$

Smooth p-Wasserstein vs Sobolev IPM: Comparison

Theorem (Dolbeault-Nazaret-Savaré'09)

Let $P, Q \in \mathcal{P}_p(\mathbb{R}^d)$ *satisfy* $P, Q \ll \gamma$ *with* $\frac{\mathsf{d}P}{\mathsf{d}\gamma} \geq c > 0$. *Then*

$$\mathsf{W}_p(P,Q) \leq c^{-1/q} p \, \mathsf{d}_{\gamma,p}(P,Q).$$

Comments:

1. Proof via Benamou-Brenier dynamical formulation

$$\mathsf{W}_p(P,Q) = \inf \left\{ \int_0^1 \|v_t\|_{L^p(\mu_t)} \, \mathsf{d}t : \mu_0 = P, \mu_1 = Q, \; \partial_t \mu_t + \nabla \cdot (v_t \mu_t) = 0 \right\}$$

   - Take $\mu_t = (1-t)P + tQ$

2. Lower bound automatically holds after Gaussian smoothing w/ $\gamma = \mathcal{N}_\sigma$

Theorem (Nietert-ZG-Kato'21)

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

$$\mathsf{W}_p^{(\sigma)}(P,Q) \leq p \, e^{\mathbb{E}[\|X\|^2]/(2q\sigma^2)} \mathsf{d}_p^{(\sigma)}(P,Q).$$

Smooth p-Wasserstein: Fast Empirical Convergence

Theorem (Nietert-ZG-Kato'21)

For P β-sub-Gaussian with $\beta < \sigma/\sqrt{2(p-1)}$: $\mathbb{E}\Big[\mathsf{W}_p^{(\sigma)}(P_n, P)\Big] \lesssim n^{-\frac{1}{2}}$.

Smooth p-Wasserstein: Fast Empirical Convergence

Theorem (Nietert-ZG-Kato'21)

For P β-sub-Gaussian with $\beta < \sigma/\sqrt{2(p-1)}$: $\mathbb{E}\Big[\mathsf{W}_p^{(\sigma)}(P_n, P)\Big] \lesssim n^{-\frac{1}{2}}$.

Pf Idea: Hinges on comparison $\mathsf{W}_p^{(\sigma)}(P,Q) \lesssim_{p,\sigma} \mathsf{d}_p^{(\sigma)}(P,Q)$

Smooth p-Wasserstein: Fast Empirical Convergence

Theorem (Nietert-ZG-Kato'21)

For P β-sub-Gaussian with $\beta < \sigma/\sqrt{2(p-1)}$: $\mathbb{E}\left[\mathsf{W}_p^{(\sigma)}(P_n, P)\right] \lesssim n^{-\frac{1}{2}}$.

Pf Idea: Hinges on comparison $\mathsf{W}_p^{(\sigma)}(P,Q) \lesssim_{p,\sigma} \mathsf{d}_p^{(\sigma)}(P,Q)$

1. Control derivative of smooth Sobolev class

Smooth p-Wasserstein: Fast Empirical Convergence

Theorem (Nietert-ZG-Kato'21)

For P β-sub-Gaussian with $\beta < \sigma/\sqrt{2(p-1)}$: $\mathbb{E}\Big[\mathsf{W}_p^{(\sigma)}(P_n, P)\Big] \lesssim n^{-\frac{1}{2}}$.

Pf Idea: Hinges on comparison $\mathsf{W}_p^{(\sigma)}(P,Q) \lesssim_{p,\sigma} \mathsf{d}_p^{(\sigma)}(P,Q)$

1. Control derivative of smooth Sobolev class

   $\Longrightarrow$ Class is P-Donsker (needs the sub-Gaussian assumption)

Smooth p-Wasserstein: Fast Empirical Convergence

Theorem (Nietert-ZG-Kato'21)

For P β-sub-Gaussian with $\beta < \sigma/\sqrt{2(p-1)}$: $\mathbb{E}\left[\mathsf{W}_p^{(\sigma)}(P_n, P)\right] \lesssim n^{-\frac{1}{2}}$.

Pf Idea: Hinges on comparison $\mathsf{W}_p^{(\sigma)}(P,Q) \lesssim_{p,\sigma} \mathsf{d}_p^{(\sigma)}(P,Q)$

1. Control derivative of smooth Sobolev class

   $\implies$ Class is P-Donsker (needs the sub-Gaussian assumption)

2. Derive limit distribution for $\sqrt{n}\mathsf{d}_p^{(\sigma)}(P_n, P)$

Smooth p-Wasserstein: Fast Empirical Convergence

Theorem (Nietert-ZG-Kato'21)

For P β-sub-Gaussian with $\beta < \sigma/\sqrt{2(p-1)}$: $\mathbb{E}\left[\mathsf{W}_p^{(\sigma)}(P_n, P)\right] \lesssim n^{-\frac{1}{2}}$.

Pf Idea: Hinges on comparison $\mathsf{W}_p^{(\sigma)}(P,Q) \lesssim_{p,\sigma} \mathsf{d}_p^{(\sigma)}(P,Q)$

1. Control derivative of smooth Sobolev class

   $\Longrightarrow$ Class is P-Donsker (needs the sub-Gaussian assumption)

2. Derive limit distribution for $\sqrt{n}\mathsf{d}_p^{(\sigma)}(P_n, P)$

3. Use comparison to obtain fast empirical convergence

Smooth p-Wasserstein: Fast Empirical Convergence

Theorem (Nietert-ZG-Kato'21)

For P β-sub-Gaussian with $\beta < \sigma/\sqrt{2(p-1)}$: $\mathbb{E}\Big[\mathsf{W}_p^{(\sigma)}(P_n, P)\Big] \lesssim n^{-\frac{1}{2}}$.

Pf Idea: Hinges on comparison $\mathsf{W}_p^{(\sigma)}(P,Q) \lesssim_{p,\sigma} \mathsf{d}_p^{(\sigma)}(P,Q)$

1. Control derivative of smooth Sobolev class

   $\implies$ Class is P-Donsker (needs the sub-Gaussian assumption)

2. Derive limit distribution for $\sqrt{n}\mathsf{d}_p^{(\sigma)}(P_n, P)$

3. Use comparison to obtain fast empirical convergence

All structural and statistical properties of $\mathsf{W}_1^{(\sigma)}$ transfer to $\mathsf{W}_p^{(\sigma)}$

Smooth Distances for Generative Modeling

Smooth Generative Models: MDE wrt smooth distance

Smooth Distances for Generative Modeling

Smooth Generative Models: MDE wrt smooth distance

1. **Generalization:** $\mathsf{W}_p^{(\sigma)}(P, Q_{\hat{\theta}_n}) - \inf_\theta \mathsf{W}_p^{(\sigma)}(P, Q_\theta) \lesssim n^{-\frac{1}{2}}, \quad \forall d$

Smooth Distances for Generative Modeling

Smooth Generative Models: MDE wrt smooth distance

1. **Generalization:** $\mathsf{W}_p^{(\sigma)}(P, Q_{\hat{\theta}_n}) - \inf_\theta \mathsf{W}_p^{(\sigma)}(P, Q_\theta) \lesssim n^{-\frac{1}{2}}, \quad \forall d$
2. **Limit distributions:** Asymptotic dist. of MDE and empirical error

Smooth Distances for Generative Modeling

Smooth Generative Models: MDE wrt smooth distance

1. **Generalization:** $\mathsf{W}_p^{(\sigma)}(P, Q_{\hat{\theta}_n}) - \inf_\theta \mathsf{W}_p^{(\sigma)}(P, Q_\theta) \lesssim n^{-\frac{1}{2}}, \quad \forall d$
2. **Limit distributions:** Asymptotic dist. of MDE and empirical error
3. **Inequalities:** Web of relationships between smooth distances

Smooth Distances for Generative Modeling

Smooth Generative Models: MDE wrt smooth distance

1. **Generalization:** $\mathsf{W}_p^{(\sigma)}(P, Q_{\hat{\theta}_n}) - \inf_\theta \mathsf{W}_p^{(\sigma)}(P, Q_\theta) \lesssim n^{-\frac{1}{2}}, \quad \forall d$
2. **Limit distributions:** Asymptotic dist. of MDE and empirical error
3. **Inequalities:** Web of relationships between smooth distances

$\Longrightarrow$ Compatible for high-dimensional learning and inference

Smooth Distances for Generative Modeling

Smooth Generative Models: MDE wrt smooth distance

1. **Generalization:** $\mathsf{W}_p^{(\sigma)}(P, Q_{\hat{\theta}_n}) - \inf_\theta \mathsf{W}_p^{(\sigma)}(P, Q_\theta) \lesssim n^{-\frac{1}{2}}, \quad \forall d$
2. **Limit distributions:** Asymptotic dist. of MDE and empirical error
3. **Inequalities:** Web of relationships between smooth distances

$\Longrightarrow$ Compatible for high-dimensional learning and inference

Future Goals: More distances, kernel, and efficient algorithms

Smooth Distances for Generative Modeling

Smooth Generative Models: MDE wrt smooth distance

1. **Generalization:** $\mathsf{W}_p^{(\sigma)}(P, Q_{\hat{\theta}_n}) - \inf_\theta \mathsf{W}_p^{(\sigma)}(P, Q_\theta) \lesssim n^{-\frac{1}{2}}, \quad \forall d$
2. **Limit distributions:** Asymptotic dist. of MDE and empirical error
3. **Inequalities:** Web of relationships between smooth distances

$\Longrightarrow$ Compatible for high-dimensional learning and inference

Future Goals: More distances, kernel, and efficient algorithms

⊛ **More distances:** f-divergences and other IPMs

Smooth Distances for Generative Modeling

Smooth Generative Models: MDE wrt smooth distance

1. **Generalization:** $\mathsf{W}_p^{(\sigma)}(P, Q_{\hat{\theta}_n}) - \inf_\theta \mathsf{W}_p^{(\sigma)}(P, Q_\theta) \lesssim n^{-\frac{1}{2}}, \quad \forall d$
2. **Limit distributions:** Asymptotic dist. of MDE and empirical error
3. **Inequalities:** Web of relationships between smooth distances

$\Longrightarrow$ Compatible for high-dimensional learning and inference

Future Goals: More distances, kernel, and efficient algorithms

- **More distances:** f-divergences and other IPMs
- **More kernels:** Optimize over choice of smoothing kernel

Smooth Distances for Generative Modeling

Smooth Generative Models: MDE wrt smooth distance

1. **Generalization:** $\mathsf{W}_p^{(\sigma)}(P, Q_{\hat{\theta}_n}) - \inf_\theta \mathsf{W}_p^{(\sigma)}(P, Q_\theta) \lesssim n^{-\frac{1}{2}}, \quad \forall d$
2. **Limit distributions:** Asymptotic dist. of MDE and empirical error
3. **Inequalities:** Web of relationships between smooth distances

$\implies$ Compatible for high-dimensional learning and inference

Future Goals: More distances, kernel, and efficient algorithms

- **More distances:** f-divergences and other IPMs
- **More kernels:** Optimize over choice of smoothing kernel
- **Efficient algorithms:** Fast computational methods

⋮

Smooth Distances for Generative Modeling

Smooth Generative Models: MDE wrt smooth distance

1. **Generalization:** $\mathsf{W}_p^{(\sigma)}(P, Q_{\hat{\theta}_n}) - \inf_\theta \mathsf{W}_p^{(\sigma)}(P, Q_\theta) \lesssim n^{-\frac{1}{2}}, \quad \forall d$
2. **Limit distributions:** Asymptotic dist. of MDE and empirical error
3. **Inequalities:** Web of relationships between smooth distances

$\Longrightarrow$ Compatible for high-dimensional learning and inference

Future Goals: More distances, kernel, and efficient algorithms

- **More distances:** f-divergences and other IPMs
- **More kernels:** Optimize over choice of smoothing kernel
- **Efficient algorithms:** Fast computational methods

⋮

Global framework for high-dimensional inference rooted in theory

Summary

- **Classic $\mathbf{W}_{\boldsymbol{p}}$:** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ with lots of applications

Summary

- **Classic $\mathbf{W}_p$:** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ with lots of applications
  - W-GAN for generative modeling ($p = 1$)

Summary

- **Classic $\mathbf{W}_{\boldsymbol{p}}$:** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ with lots of applications
  - W-GAN for generative modeling ($p = 1$)
  - Empirical approximation is slow $n^{-1/d}$

Summary

- **Classic W_p:** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ with lots of applications
  - W-GAN for generative modeling ($p = 1$)
  - Empirical approximation is slow $n^{-1/d}$
- **Smooth $\mathsf{W}_p^{(\sigma)}$:** Convolve distributions w/ Gaussian kernel

Summary

- **Classic W_p:** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ with lots of applications
  - W-GAN for generative modeling ($p = 1$)
  - Empirical approximation is slow $n^{-1/d}$
- **Smooth $\mathsf{W}_p^{(\sigma)}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_p **&** duality

Summary

- **Classic $\mathbf{W}_{p}$:** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ with lots of applications
  - W-GAN for generative modeling ($p = 1$)
  - Empirical approximation is slow $n^{-1/d}$
- **Smooth $\mathbf{W}_{p}^{(\sigma)}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_p **&** duality
  - Well-behaved function of smoothing parameter **&** recovers W_p in limit

Summary

- **Classic W_p:** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ with lots of applications
  - W-GAN for generative modeling ($p = 1$)
  - Empirical approximation is slow $n^{-1/d}$
- **Smooth $\mathsf{W}_p^{(\sigma)}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_p **&** duality
  - Well-behaved function of smoothing parameter **&** recovers W_p in limit
  - Fast $n^{-1/2}$ empirical convergence in all dimensions

Summary

- **Classic $\mathbf{W}_p$:** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ with lots of applications
  - W-GAN for generative modeling ($p = 1$)
  - Empirical approximation is slow $n^{-1/d}$
- **Smooth $\mathbf{W}_p^{(\sigma)}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_p **&** duality
  - Well-behaved function of smoothing parameter **&** recovers W_p in limit
  - Fast $n^{-1/2}$ empirical convergence in all dimensions
  - Limit distribution for $\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P)$ in all dimensions

Summary

- **Classic W_p:** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ with lots of applications
  - W-GAN for generative modeling ($p = 1$)
  - Empirical approximation is slow $n^{-1/d}$
- **Smooth $\mathsf{W}_p^{(\sigma)}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_p **&** duality
  - Well-behaved function of smoothing parameter **&** recovers W_p in limit
  - Fast $n^{-1/2}$ empirical convergence in all dimensions
  - Limit distribution for $\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P)$ in all dimensions
- **Smooth MDE:** Generative modeling w/ performance guarantees

Summary

- **Classic W_p:** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ with lots of applications
  - W-GAN for generative modeling ($p = 1$)
  - Empirical approximation is slow $n^{-1/d}$
- **Smooth $\mathsf{W}_p^{(\sigma)}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_p **&** duality
  - Well-behaved function of smoothing parameter **&** recovers W_p in limit
  - Fast $n^{-1/2}$ empirical convergence in all dimensions
  - Limit distribution for $\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P)$ in all dimensions
- **Smooth MDE:** Generative modeling w/ performance guarantees
  - Attain competitive performance

Summary

- **Classic W_p:** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ with lots of applications
  - W-GAN for generative modeling ($p = 1$)
  - Empirical approximation is slow $n^{-1/d}$
- **Smooth $\mathsf{W}_p^{(\sigma)}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_p **&** duality
  - Well-behaved function of smoothing parameter **&** recovers W_p in limit
  - Fast $n^{-1/2}$ empirical convergence in all dimensions
  - Limit distribution for $\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P)$ in all dimensions
- **Smooth MDE:** Generative modeling w/ performance guarantees
  - Attain competitive performance
  - Consume fewer resources

Summary

- **Classic W_p:** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ with lots of applications
  - W-GAN for generative modeling ($p = 1$)
  - Empirical approximation is slow $n^{-1/d}$
- **Smooth $\mathsf{W}_p^{(\sigma)}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_p **&** duality
  - Well-behaved function of smoothing parameter **&** recovers W_p in limit
  - Fast $n^{-1/2}$ empirical convergence in all dimensions
  - Limit distribution for $\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P)$ in all dimensions
- **Smooth MDE:** Generative modeling w/ performance guarantees
  - Attain competitive performance
  - Consume fewer resources
  - Backed up by (almost useful) generalization bounds

Summary

- **Classic W_p:** Metric on $\mathcal{P}_p(\mathbb{R}^d)$ with lots of applications
  - W-GAN for generative modeling ($p = 1$)
  - Empirical approximation is slow $n^{-1/d}$
- **Smooth $\mathsf{W}_p^{(\sigma)}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_p **&** duality
  - Well-behaved function of smoothing parameter **&** recovers W_p in limit
  - Fast $n^{-1/2}$ empirical convergence in all dimensions
  - Limit distribution for $\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P)$ in all dimensions
- **Smooth MDE:** Generative modeling w/ performance guarantees
  - Attain competitive performance
  - Consume fewer resources
  - Backed up by (almost useful) generalization bounds

Thank you!