Smooth Wasserstein Distance: Metric Structure and Statistical Efficiency

Ziv Goldfeld

Cornell University

July 2020

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

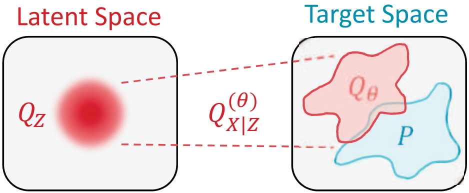
Implicit (Latent Variable) Generative Models

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



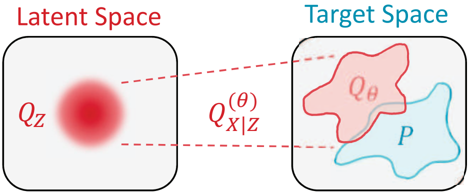
Implicit (Latent Variable) Generative Models

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 $\theta^\star \in \underset{\theta}{\operatorname{argmin}}\, \delta(P, Q_\theta)$

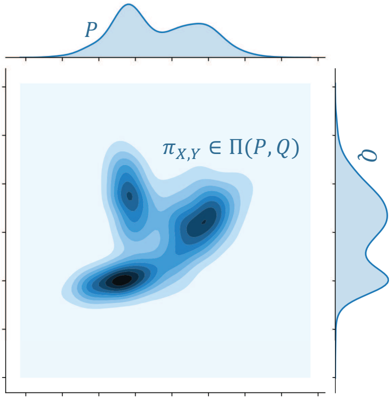
The 1-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_1(\mathbb{R}^d)$ (subscript for finite 1st moments)

The 1-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_1(\mathbb{R}^d)$ (subscript for finite 1st moments)

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



The 1-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_1(\mathbb{R}^d)$ (subscript for finite 1st moments)

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

The 1-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_1(\mathbb{R}^d)$ (subscript for finite 1st moments)

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

Definition (1-Wasserstein)

The 1-Wasserstein distance: $\mathsf{W}_1(P,Q) := \inf_{\pi_{X,Y} \in \Pi(P,Q)} \mathbb{E}_\pi \|X - Y\|$

The 1-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_1(\mathbb{R}^d)$ (subscript for finite 1st moments)

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

Definition (1-Wasserstein)

The 1-Wasserstein distance: $\mathsf{W}_1(P,Q) := \inf_{\pi_{X,Y} \in \Pi(P,Q)} \mathbb{E}_\pi \|X - Y\|$

Comments:

The 1-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_1(\mathbb{R}^d)$ (subscript for finite 1st moments)

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

Definition (1-Wasserstein)

The 1-Wasserstein distance: $\mathsf{W}_1(P,Q) := \inf_{\pi_{X,Y} \in \Pi(P,Q)} \mathbb{E}_\pi \|X-Y\|$

Comments:

- **Robustness to Supp. Mismatch:** $\mathsf{W}_1(P,Q) < \infty$, $\forall P, Q \in \mathcal{P}_1(\mathbb{R}^d)$

The 1-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_1(\mathbb{R}^d)$ (subscript for finite 1st moments)

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

Definition (1-Wasserstein)

The 1-Wasserstein distance: $\mathsf{W}_1(P,Q) := \inf_{\pi_{X,Y} \in \Pi(P,Q)} \mathbb{E}_\pi \|X - Y\|$

Comments:

- **Robustness to Supp. Mismatch:** $\mathsf{W}_1(P,Q) < \infty$, $\forall P, Q \in \mathcal{P}_1(\mathbb{R}^d)$
- **Metric:** $\left(\mathcal{P}_1(\mathbb{R}^d), \mathsf{W}_1\right)$ is metric space (metrizes weak convergence)

The 1-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_1(\mathbb{R}^d)$ (subscript for finite 1st moments)

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

Definition (1-Wasserstein)

The 1-Wasserstein distance: $\mathsf{W}_1(P,Q) := \inf_{\pi_{X,Y} \in \Pi(P,Q)} \mathbb{E}_\pi \|X-Y\|$

Comments:

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

The 1-Wasserstein Distance

Setup: $P, Q \in \mathcal{P}_1(\mathbb{R}^d)$ (subscript for finite 1st moments)

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

Definition (1-Wasserstein)

The 1-Wasserstein distance: $\mathsf{W}_1(P,Q) := \inf_{\pi_{X,Y} \in \Pi(P,Q)} \mathbb{E}_\pi \|X - Y\|$

Comments:

- **Robustness to Supp. Mismatch:** $\mathsf{W}_1(P,Q) < \infty$, $\forall P, Q \in \mathcal{P}_1(\mathbb{R}^d)$
- **Metric:** $\big(\mathcal{P}_1(\mathbb{R}^d), \mathsf{W}_1\big)$ 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)

Implicit Generative Models: Generalization

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

Implicit Generative Models: Generalization

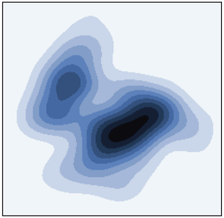
Goal: Solve $\mathsf{OPT} := \inf_{\theta} \mathsf{W}_1(P, Q_{\theta})$ exactly (find $\theta^{\star}$)

Estimation: We don't have P but data

Implicit Generative Models: Generalization

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

Estimation: We don't have P but data

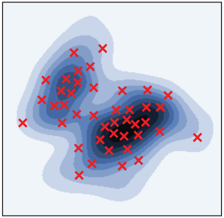


- $\{X_i\}_{i=1}^n$ are i.i.d. samples from $P \in \mathcal{P}(\mathbb{R}^d)$

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



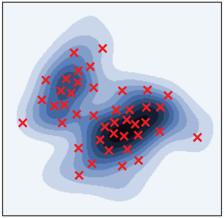
- $\{X_i\}_{i=1}^n$ are i.i.d. samples from $P \in \mathcal{P}(\mathbb{R}^d)$
- Empirical distribution $P_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$

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 $P_n := \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$



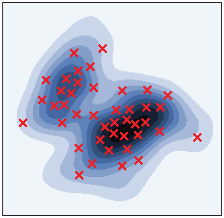
$\Longrightarrow$ Inherently we work with $\mathsf{W}_1(P_n, Q_\theta)$

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 $P_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$



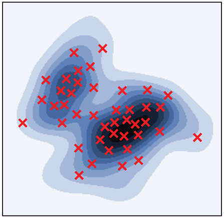
$\implies$ Inherently we work with $\mathsf{W}_1(P_n, Q_\theta)$

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 $P_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$

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

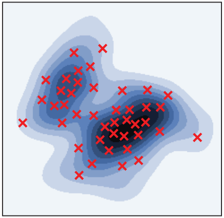
Optimization: Can solve $\inf_\theta \mathsf{W}_1(P_n, Q_\theta)$ approximately

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

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

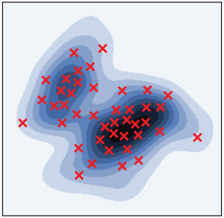
Generalization [Zhang *et al*'18]: $\mathsf{W}_1(P, Q_{\hat{\theta}_n}) -$ OPT $\leq 2\mathsf{W}_1(P_n, P) + \epsilon$

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 $P_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$



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

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

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 [Zhang *et al*'18]: $\mathsf{W}_1(P, Q_{\hat{\theta}_n}) - \text{OPT} \leq 2\mathsf{W}_1(P_n, P) + \epsilon$

$\implies$ **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* $\mathcal{P}_1(\mathbb{R}^d) \ni P \ll \mathsf{Leb}(\mathbb{R}^d)$*:* $\mathbb{E}\mathsf{W}_1(P_n, P) \asymp n^{-\frac{1}{d}}$

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* $\mathcal{P}_1(\mathbb{R}^d) \ni P \ll \mathsf{Leb}(\mathbb{R}^d)$*:* $\mathbb{E}\mathsf{W}_1(P_n, P) \asymp n^{-\frac{1}{d}}$

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* $\mathcal{P}_1(\mathbb{R}^d) \ni P \ll \mathsf{Leb}(\mathbb{R}^d)$: $\mathbb{E}\mathsf{W}_1(P_n, P) \asymp n^{-\frac{1}{d}}$

⊛ **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 $\mathcal{P}_1(\mathbb{R}^d) \ni P \ll \mathrm{Leb}(\mathbb{R}^d)$: $\mathbb{E}\mathsf{W}_1(P_n, P) \asymp n^{-\frac{1}{d}}$

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

⊛ **Question:** Can we find a new W_1-like distance that alleviates CoD?

Smooth 1-Wasserstein Distance

Definition (ZG-Greenewald'19)

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

Smooth 1-Wasserstein Distance

Definition (ZG-Greenewald'19)

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 $\mathcal{N}_\sigma \triangleq \mathcal{N}(0, \sigma^2 \mathrm{I}_d)$ 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 (ZG-Greenewald'19)

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 $\mathcal{N}_\sigma \triangleq \mathcal{N}(0, \sigma^2 \mathrm{I}_d)$ 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 (ZG-Greenewald'19)

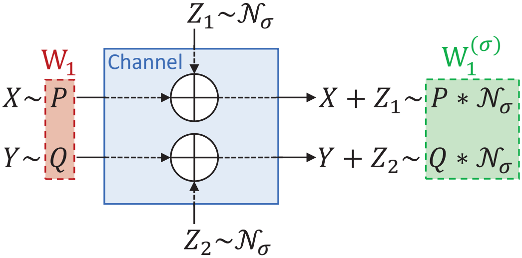
For $\sigma \geq 0$, the smooth 1-Wasserstein distance between P and Q is

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

where $\mathcal{N}_\sigma \triangleq \mathcal{N}(0, \sigma^2 \mathrm{I}_d)$ 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 (ZG-Greenewald'19)

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

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

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

Retain 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 Distance

Definition (ZG-Greenewald'19)

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

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

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

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

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

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'19)

$\left(\mathcal{P}_1(\mathbb{R}^d), \mathsf{W}_1^{(\sigma)}\right)$ *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'19)

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

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

Smooth 1-Wasserstein – Metric Structure

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

Theorem (ZG-Greenewald'19)

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

Key Idea for Pf.: Use characteristic functions $\Phi_P(t) := \mathbb{E}_P\left[e^{itX}\right]$ 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'19)

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

Key Idea for Pf.: Use characteristic functions $\Phi_P(t) := \mathbb{E}_P\left[e^{itX}\right]$ 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'19)

Let $P_n, P \in \mathcal{P}_1(\mathbb{R}^d)$, $n \geq 1$. *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'19)

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

Key Idea for Pf.: Use characteristic functions $\Phi_P(t) := \mathbb{E}_P\left[e^{itX}\right]$ 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'19)

Let $P_n, P \in \mathcal{P}_1(\mathbb{R}^d)$, $n \geq 1$. *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'19)

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'19)

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'19)

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'19)

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'19)

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'19)

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'19)

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

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'19)

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'19)

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

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 – 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'19)

Fix $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$ *and* $\sigma_k \searrow 0$. *Let:*

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'19)

Fix $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$ *and* $\sigma_k \searrow 0$. *Let:*

- $\{\pi_k\}$, *s.t.* $\pi_k \in \Pi(P * \mathcal{N}_{\sigma_k}, Q * \mathcal{N}_{\sigma_k})$ *is optimal for* $\mathsf{W}_1^{(\sigma_k)}(P,Q)$

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'19)

Fix $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$ *and* $\sigma_k \searrow 0$. *Let:*

- $\{\pi_k\}$, *s.t.* $\pi_k \in \Pi(P*\mathcal{N}_{\sigma_k}, Q*\mathcal{N}_{\sigma_k})$ *is optimal for* $\mathsf{W}_1^{(\sigma_k)}(P,Q)$

Then $\pi_k \xrightarrow[k\to\infty]{} \pi$ *(weakly) such that:*

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'19)

Fix $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$ *and* $\sigma_k \searrow 0$*. Let:*

- $\{\pi_k\}$, *s.t.* $\pi_k \in \Pi(P * \mathcal{N}_{\sigma_k}, Q * \mathcal{N}_{\sigma_k})$ *is optimal for* $\mathsf{W}_1^{(\sigma_k)}(P,Q)$

Then $\pi_k \xrightarrow[k\to\infty]{} \pi$ *(weakly) such that:*

- $\pi \in \Pi(P,Q)$

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'19)

Fix $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$ *and* $\sigma_k \searrow 0$. *Let:*

- $\{\pi_k\}$, *s.t.* $\pi_k \in \Pi(P * \mathcal{N}_{\sigma_k}, Q * \mathcal{N}_{\sigma_k})$ *is optimal for* $\mathsf{W}_1^{(\sigma_k)}(P,Q)$

Then $\pi_k \xrightarrow[k\to\infty]{} \pi$ *(weakly) such that:*

- $\pi \in \Pi(P,Q)$
- π *is optimal for* $\mathsf{W}_1(P,Q)$

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'19)

Fix $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$ *and* $\sigma_k \searrow 0$. *Let:*

- $\{\pi_k\}$, *s.t.* $\pi_k \in \Pi(P * \mathcal{N}_{\sigma_k}, Q * \mathcal{N}_{\sigma_k})$ *is optimal for* $\mathsf{W}_1^{(\sigma_k)}(P,Q)$

Then $\pi_k \xrightarrow[k\to\infty]{} \pi$ *(weakly) such that:*

- $\pi \in \Pi(P,Q)$
- π *is optimal for* $\mathsf{W}_1(P,Q)$

Comments:

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'19)

Fix $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$ *and* $\sigma_k \searrow 0$. *Let:*

- $\{\pi_k\}$, *s.t.* $\pi_k \in \Pi(P*\mathcal{N}_{\sigma_k}, Q*\mathcal{N}_{\sigma_k})$ *is optimal for* $\mathsf{W}_1^{(\sigma_k)}(P,Q)$

Then $\pi_k \xrightarrow[k\to\infty]{} \pi$ *(weakly) such that:*

- $\pi \in \Pi(P,Q)$
- π *is optimal for* $\mathsf{W}_1(P,Q)$

Comments:

- **In words:** Not only opt. values converge, but also optimizers

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'19)

Fix $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$ *and* $\sigma_k \searrow 0$. *Let:*

- $\{\pi_k\}$, *s.t.* $\pi_k \in \Pi(P * \mathcal{N}_{\sigma_k}, Q * \mathcal{N}_{\sigma_k})$ *is optimal for* $\mathsf{W}_1^{(\sigma_k)}(P,Q)$

Then $\pi_k \xrightarrow[k\to\infty]{} \pi$ *(weakly) such that:*

- $\pi \in \Pi(P,Q)$
- π *is optimal for* $\mathsf{W}_1(P,Q)$

Comments:

- **In words:** Not only opt. values converge, but also optimizers
- **Generalization:** Also holds for $\sigma_k \searrow \sigma > 0$

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'19)

Fix $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$ and $\sigma_k \searrow 0$. Let:

- *$\{\pi_k\}$, s.t. $\pi_k \in \Pi(P * \mathcal{N}_{\sigma_k}, Q * \mathcal{N}_{\sigma_k})$ is optimal for $\mathsf{W}_1^{(\sigma_k)}(P,Q)$*

Then $\pi_k \xrightarrow[k\to\infty]{} \pi$ (weakly) such that:

- $\pi \in \Pi(P,Q)$
- *π is optimal for $\mathsf{W}_1(P,Q)$*

Comments:

- **In words:** Not only opt. values converge, but also optimizers
- **Generalization:** Also holds for $\sigma_k \searrow \sigma > 0$
- **Pf. Idea:** Γ-convergence & Tightness of $\Pi(P,Q)$

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'19)

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

Smooth 1-Wasserstein – Statistical Efficiency

High Level: Alleviate curse of dimensionality & get limit distribution

Theorem (ZG-Greenewald-Polyanskiy-Weed'19)

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

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'19)

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

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 $n^{-\frac{1}{2}}$ rate is sharp

Smooth 1-Wasserstein – Statistical Efficiency

High Level: Alleviate curse of dimensionality & get limit distribution

Theorem (ZG-Greenewald-Polyanskiy-Weed'19)

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

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 $n^{-\frac{1}{2}}$ 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'19)

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

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 $n^{-\frac{1}{2}}$ 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 in all d

Smooth 1-Wasserstein – Statistical Efficiency

High Level: Alleviate curse of dimensionality & get limit distribution

Theorem (ZG-Greenewald-Polyanskiy-Weed'19)

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

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 $n^{-\frac{1}{2}}$ 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 in all $d \implies O(n^{-\frac{1}{2}})$ GAN Generalization

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

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

Limit Distribution Proof Roadmap

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

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

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

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

⊛ **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 $\mathcal{F}_\sigma := \big\{ f * \varphi_\sigma : \; f \in \mathsf{Lip}_1(\mathbb{R}^d) \big\}$ 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. $\mathbf{W}_1^{(\sigma)}$:**

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

⊛ **Objective:** $\mathbb{G}_n = \big(\sqrt{n}(P_n - P)g\big)_{g \in \mathcal{F}_\sigma} \xrightarrow{w} \mathsf{GP} \iff \mathcal{F}_\sigma$ 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. $\mathbf{W}_1^{(\sigma)}$:**

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

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

2. **[van der Vaart'96]:** $K \subseteq \mathbb{R}^d$ bdd & convex: $\mathcal{F}_\sigma\Big|_{\mathcal{K}} \subseteq C_M^N(\mathcal{K}),\ N > \frac{d}{2}$

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

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

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

2. **[van der Vaart'96]:** $K \subseteq \mathbb{R}^d$ bdd & convex: $\mathcal{F}_\sigma\Big|_{\mathcal{K}} \subseteq C_M^N(\mathcal{K}), \ N > \frac{d}{2}$

$\Longrightarrow$ **$\mathcal{F}_\sigma$ is P-Donsker:** $\boxed{\mathbb{G}_n \stackrel{w}{\to} G_P \text{ (Gaussian process)}}$

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

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

⊛ **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. **[van der Vaart'96]:** $K \subseteq \mathbb{R}^d$ bdd & convex: $\mathcal{F}_\sigma\Big|_{\mathcal{K}} \subseteq C_M^N(\mathcal{K}),\ N > \frac{d}{2}$

$\Longrightarrow$ **$\mathcal{F}_\sigma$ is P-Donsker:** $\mathbb{G}_n \xrightarrow{w} G_P$ (Gaussian process)

3. **Continuous Mapping Theorem:** $\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P) \xrightarrow{d} \sup_{g \in \mathcal{F}_\sigma} G_P(g)$

Generative Models: Generalization & Computation

Generalization: Want to solve $\mathsf{OPT} := \inf_{\theta} \delta\left(P, Q_{\theta}\right)$ exactly (find $\theta^{\star}$)

Generative Models: Generalization & Computation

<u>**Generalization:**</u> Want to solve OPT $:= \inf_\theta \delta(P, Q_\theta)$ exactly (find $\theta^\star$)

- **In practice:** Can solve $\inf_\theta \delta(P_n, Q_\theta)$ approximately:

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

Generative Models: Generalization & Computation

Generalization: Want to solve OPT $:= \inf_\theta \delta\left(P, Q_\theta\right)$ exactly (find $\theta^\star$)

- **In practice:** Can solve $\inf_\theta \delta\left(P_n, Q_\theta\right)$ approximately:

$$\text{Find } \hat{\theta}_n \text{ s.t. } \delta\big(P_n, Q_{\hat{\theta}_n}\big) \leq \inf_\theta \delta\left(P_n, Q_\theta\right) + \epsilon$$

- **Guarantees:** For $\delta = \mathsf{W}_1^{(\sigma)}$ and $\epsilon = n^{-1/2}$

Generative Models: Generalization & Computation

Generalization: Want to solve OPT $:= \inf_\theta \delta(P, Q_\theta)$ exactly (find $\theta^\star$)

- **In practice:** Can solve $\inf_\theta \delta(P_n, Q_\theta)$ approximately:

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

- **Guarantees:** For $\delta = \mathsf{W}_1^{(\sigma)}$ and $\epsilon = n^{-1/2}$

[Zhang *et al*'18]: $\mathsf{W}_1^{(\sigma)}\big(P, Q_{\hat{\theta}_n}\big) - \mathrm{OPT} \leq 2\mathsf{W}_1^{(\sigma)}(P_n, P) + n^{-\frac{1}{2}}$

Generative Models: Generalization & Computation

Generalization: Want to solve OPT $:= \inf_\theta \delta(P, Q_\theta)$ exactly (find $\theta^\star$)

- **In practice:** Can solve $\inf_\theta \delta(P_n, Q_\theta)$ approximately:

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

- **Guarantees:** For $\delta = \mathsf{W}_1^{(\sigma)}$ and $\epsilon = n^{-1/2}$

[Zhang *et al*'18]: $\mathsf{W}_1^{(\sigma)}\big(P, Q_{\hat{\theta}_n}\big) - \mathsf{OPT} \le 2\mathsf{W}_1^{(\sigma)}(P_n, P) + n^{-\frac{1}{2}} \lesssim n^{-\frac{1}{2}}$

Generative Models: Generalization & Computation

Generalization: Want to solve OPT $:= \inf_\theta \delta(P, Q_\theta)$ exactly (find $\theta^\star$)

- **In practice:** Can solve $\inf_\theta \delta(P_n, Q_\theta)$ approximately:

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

- **Guarantees:** For $\delta = \mathsf{W}_1^{(\sigma)}$ and $\epsilon = n^{-1/2}$

[Zhang *et al*'18]: $\mathsf{W}_1^{(\sigma)}(P, Q_{\hat{\theta}_n}) - \mathsf{OPT} \le 2\mathsf{W}_1^{(\sigma)}(P_n, P) + n^{-\frac{1}{2}} \lesssim n^{-\frac{1}{2}}$

[ZG-Kato'20]: $\mathbb{E}\big[\hat{\theta}_n - \theta^\star\big] \asymp n^{-\frac{1}{2}}$ (in fact, $\sqrt{n}(\hat{\theta}_n - \theta^\star)$ converges in dist.)

Generative Models: Generalization & Computation

Generalization: Want to solve OPT $:= \inf_\theta \delta(P, Q_\theta)$ exactly (find $\theta^\star$)

- **In practice:** Can solve $\inf_\theta \delta(P_n, Q_\theta)$ approximately:

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

- **Guarantees:** For $\delta = \mathsf{W}_1^{(\sigma)}$ and $\epsilon = n^{-1/2}$

[Zhang *et al*'18]: $\mathsf{W}_1^{(\sigma)}\big(P, Q_{\hat{\theta}_n}\big) - \mathsf{OPT} \leq 2\mathsf{W}_1^{(\sigma)}(P_n, P) + n^{-\frac{1}{2}} \lesssim n^{-\frac{1}{2}}$

[ZG-Kato'20]: $\mathbb{E}\big[\hat{\theta}_n - \theta^\star\big] \asymp n^{-\frac{1}{2}}$ (in fact, $\sqrt{n}(\hat{\theta}_n - \theta^\star)$ converges in dist.)

Computation: How to compute $\mathsf{W}_1^{(\sigma)}(P, Q)$?

Generative Models: Generalization & Computation

Generalization: Want to solve OPT $:= \inf_\theta \delta(P, Q_\theta)$ exactly (find $\theta^\star$)

- **In practice:** Can solve $\inf_\theta \delta(P_n, Q_\theta)$ approximately:

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

- **Guarantees:** For $\delta = \mathsf{W}_1^{(\sigma)}$ and $\epsilon = n^{-1/2}$

  [Zhang *et al*'18]: $\mathsf{W}_1^{(\sigma)}\big(P, Q_{\hat{\theta}_n}\big) - \mathsf{OPT} \le 2\mathsf{W}_1^{(\sigma)}(P_n, P) + n^{-\frac{1}{2}} \lesssim n^{-\frac{1}{2}}$

  [ZG-Kato'20]: $\mathbb{E}\big[\hat{\theta}_n - \theta^\star\big] \asymp n^{-\frac{1}{2}}$ (in fact, $\sqrt{n}(\hat{\theta}_n - \theta^\star)$ converges in dist.)

Computation: How to compute $\mathsf{W}_1^{(\sigma)}(P, Q)$?

- **Naive:** $\sup_{f \in \mathsf{Lip}_1} \mathbb{E}f(X+Z) - \mathbb{E}f(Y+Z)$ & MC sampling for Z

Generative Models: Generalization & Computation

Generalization: Want to solve OPT $:= \inf_\theta \delta(P, Q_\theta)$ exactly (find $\theta^\star$)

- **In practice:** Can solve $\inf_\theta \delta(P_n, Q_\theta)$ approximately:

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

- **Guarantees:** For $\delta = \mathsf{W}_1^{(\sigma)}$ and $\epsilon = n^{-1/2}$

  [Zhang *et al*'18]: $\mathsf{W}_1^{(\sigma)}\big(P, Q_{\hat{\theta}_n}\big) - \text{OPT} \leq 2\mathsf{W}_1^{(\sigma)}(P_n, P) + n^{-\frac{1}{2}} \lesssim n^{-\frac{1}{2}}$

  [ZG-Kato'20]: $\mathbb{E}\big[\hat{\theta}_n - \theta^\star\big] \asymp n^{-\frac{1}{2}}$ (in fact, $\sqrt{n}(\hat{\theta}_n - \theta^\star)$ converges in dist.)

Computation: How to compute $\mathsf{W}_1^{(\sigma)}(P, Q)$?

- **Naive:** $\sup_{f \in \mathsf{Lip}_1} \mathbb{E} f(X+Z) - \mathbb{E} f(Y+Z)$ & MC sampling for Z
- **Advanced:** $\sup_{f \in \mathsf{Lip}_1} \mathbb{E}(f * \varphi_\sigma)(X) - \mathbb{E}(f * \varphi_\sigma)(Y)$ & implement '$*$'

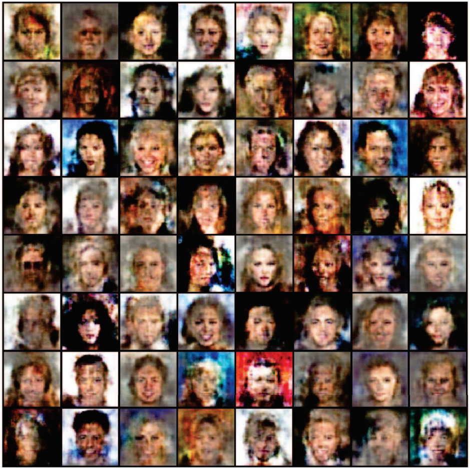
Smooth Wasserstein GAN: Initial Empirical Results

Setup: f parametrized by 2-layer NN & conv. computed in closed form

Smooth Wasserstein GAN: Initial Empirical Results

Setup: f parametrized by 2-layer NN $\&$ conv. computed in closed form

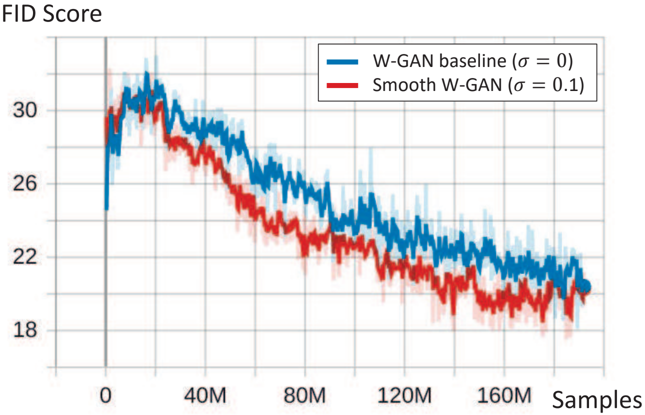
Baseline W-GAN ($\sigma = 0$)



Smooth W-GAN ($\sigma = 0.1$)

Smooth Wasserstein GAN: Initial Empirical Results

Setup: f parametrized by 2-layer NN & conv. computed in closed form



⊛ FID = Fréchet Inception Distance

Summary

- **Classic $\mathbf{W_1}$:** Metric on $\mathcal{P}_1(\mathbb{R}^d)$ that is popular in ML

Summary

- **Classic $\mathbf{W_1}$:** Metric on $\mathcal{P}_1(\mathbb{R}^d)$ that is popular in ML
  - W-GAN for generative modeling

Summary

- **Classic $\mathbf{W_1}$:** Metric on $\mathcal{P}_1(\mathbb{R}^d)$ that is popular in ML
  - W-GAN for generative modeling
  - Empirical approximation is slow $n^{-\frac{1}{d}}$

Summary

- **Classic $\mathbf{W_1}$:** Metric on $\mathcal{P}_1(\mathbb{R}^d)$ that is popular in ML
  - W-GAN for generative modeling
  - Empirical approximation is slow $n^{-\frac{1}{d}}$
- **Smooth $\mathbf{W_1^{(\sigma)}}$:** Convolve distributions w/ Gaussian kernel

Summary

- **Classic $\mathsf{W_1}$:** Metric on $\mathcal{P}_1(\mathbb{R}^d)$ that is popular in ML
  - W-GAN for generative modeling
  - Empirical approximation is slow $n^{-\frac{1}{d}}$
- **Smooth $\mathsf{W}_1^{(\sigma)}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_1 & duality

Summary

- **Classic W_1:** Metric on $\mathcal{P}_1(\mathbb{R}^d)$ that is popular in ML
  - W-GAN for generative modeling
  - Empirical approximation is slow $n^{-\frac{1}{d}}$
- **Smooth $\mathsf{W}_1^{(\sigma)}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_1 & duality
  - Well-behaved function of smoothing parameter & recovers W_1 in limit

Summary

- **Classic $\mathbf{W_1}$:** Metric on $\mathcal{P}_1(\mathbb{R}^d)$ that is popular in ML
  - W-GAN for generative modeling
  - Empirical approximation is slow $n^{-\frac{1}{d}}$
- **Smooth $\mathbf{W_1^{(\sigma)}}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_1 & duality
  - Well-behaved function of smoothing parameter & recovers W_1 in limit
  - Fast $n^{-\frac{1}{2}}$ empirical convergence in all dimensions

Summary

- **Classic $\mathbf{W_1}$:** Metric on $\mathcal{P}_1(\mathbb{R}^d)$ that is popular in ML
  - W-GAN for generative modeling
  - Empirical approximation is slow $n^{-\frac{1}{d}}$
- **Smooth $\mathbf{W_1^{(\sigma)}}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_1 & duality
  - Well-behaved function of smoothing parameter & recovers W_1 in limit
  - Fast $n^{-\frac{1}{2}}$ empirical convergence in all dimensions
  - Limit distribution for $\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P)$ in all dimensions

Summary

- **Classic $\mathbf{W_1}$:** Metric on $\mathcal{P}_1(\mathbb{R}^d)$ that is popular in ML
  - W-GAN for generative modeling
  - Empirical approximation is slow $n^{-\frac{1}{d}}$
- **Smooth $\mathbf{W_1^{(\sigma)}}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_1 & duality
  - Well-behaved function of smoothing parameter & recovers W_1 in limit
  - Fast $n^{-\frac{1}{2}}$ empirical convergence in all dimensions
  - Limit distribution for $\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P)$ in all dimensions
- **Smooth W-GAN:** Generative modeling w/ performance guarantees

Summary

- **Classic $\mathbf{W_1}$:** Metric on $\mathcal{P}_1(\mathbb{R}^d)$ that is popular in ML
  - W-GAN for generative modeling
  - Empirical approximation is slow $n^{-\frac{1}{d}}$
- **Smooth $\mathbf{W_1^{(\sigma)}}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_1 & duality
  - Well-behaved function of smoothing parameter & recovers W_1 in limit
  - Fast $n^{-\frac{1}{2}}$ empirical convergence in all dimensions
  - Limit distribution for $\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P)$ in all dimensions
- **Smooth W-GAN:** Generative modeling w/ performance guarantees
  - Attain competitive performance

Summary

- **Classic W_1:** Metric on $\mathcal{P}_1(\mathbb{R}^d)$ that is popular in ML
  - W-GAN for generative modeling
  - Empirical approximation is slow $n^{-\frac{1}{d}}$
- **Smooth $\mathsf{W}_1^{(\sigma)}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_1 & duality
  - Well-behaved function of smoothing parameter & recovers W_1 in limit
  - Fast $n^{-\frac{1}{2}}$ empirical convergence in all dimensions
  - Limit distribution for $\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P)$ in all dimensions
- **Smooth W-GAN:** Generative modeling w/ performance guarantees
  - Attain competitive performance
  - Consume fewer resources

Summary

- **Classic W_1:** Metric on $\mathcal{P}_1(\mathbb{R}^d)$ that is popular in ML
  - W-GAN for generative modeling
  - Empirical approximation is slow $n^{-\frac{1}{d}}$
- **Smooth $\mathsf{W}_1^{(\sigma)}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_1 & duality
  - Well-behaved function of smoothing parameter & recovers W_1 in limit
  - Fast $n^{-\frac{1}{2}}$ empirical convergence in all dimensions
  - Limit distribution for $\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P)$ in all dimensions
- **Smooth W-GAN:** Generative modeling w/ performance guarantees
  - Attain competitive performance
  - Consume fewer resources
  - Backed up by (almost useful) generalization bounds

Summary

- **Classic $\mathbf{W_1}$:** Metric on $\mathcal{P}_1(\mathbb{R}^d)$ that is popular in ML
  - W-GAN for generative modeling
  - Empirical approximation is slow $n^{-\frac{1}{d}}$
- **Smooth $\mathbf{W_1^{(\sigma)}}$:** Convolve distributions w/ Gaussian kernel
  - Inherits metric structure of W_1 & duality
  - Well-behaved function of smoothing parameter & recovers W_1 in limit
  - Fast $n^{-\frac{1}{2}}$ empirical convergence in all dimensions
  - Limit distribution for $\sqrt{n}\mathsf{W}_1^{(\sigma)}(P_n, P)$ in all dimensions
- **Smooth W-GAN:** Generative modeling w/ performance guarantees
  - Attain competitive performance
  - Consume fewer resources
  - Backed up by (almost useful) generalization bounds

Thank you!