Gaussian-Smoothed Optimal Transport: Metric Structure and Statistical Efficiency

Ziv Goldfeld

Cornell University

Colloquium, Center of Applied Mathematics, Cornell University

November 1st, 2019

Outline

1. Generative adversarial networks (GANs)
2. Optimal transport (OT) and Wasserstein metric
3. Entropic optimal transport
4. Gaussian-smoothed optimal transport
5. Summary

Generative Modeling - Preamble

- Which face is fake?

Generative Modeling - Preamble

- Which face is fake?
- **Unsupervised Learning:**

Generative Modeling - Preamble

- Which face is fake?

- **Unsupervised Learning:**
  - **Data:** $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ samples (no labels!)

Generative Modeling - Preamble

- Which face is fake?

- **Unsupervised Learning:**
  - **Data:** $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ samples (no labels!)
  - **Goal:** Learn underlying structure of data, e.g., $Q_X \approx P_X$

Generative Modeling - Preamble

- Which face is fake?

- **Unsupervised Learning:**
  - **Data:** $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ samples (no labels!)
  - **Goal:** Learn underlying structure of data, e.g., $Q_X \approx P_X$
  - **Examples:** Generative models, clustering, dim. reduction, etc.

Generative Modeling - Preamble

- Which face is fake?

- **Unsupervised Learning:**
  - **Data:** $\{X_i\}_{i=1}^{n} \overset{\text{iid}}{\sim} P_X$ samples (no labels!)
  - **Goal:** Learn underlying structure of data, e.g., $Q_X \approx P_X$
  - **Examples:** **Generative models**, clustering, dim. reduction, etc.

Generative Modeling - Preamble

- Which face is fake?

- **Unsupervised Learning:**
  - **Data:** $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ samples (no labels!)
  - **Goal:** Learn underlying structure of data, e.g., $Q_X \approx P_X$
  - **Examples:** **Generative models**, clustering, dim. reduction, etc.

- **GANs [Goodfellow *et al*'14]:**

Generative Modeling - Preamble

- Which face is fake?

- **Unsupervised Learning:**
  - **Data:** $\{X_i\}_{i=1}^{n} \overset{\text{iid}}{\sim} P_X$ samples (no labels!)
  - **Goal:** Learn underlying structure of data, e.g., $Q_X \approx P_X$
  - **Examples:** **Generative models**, clustering, dim. reduction, etc.

- **GANs [Goodfellow *et al*'14]:**
  - Sample-to-sample

Generative Modeling - Preamble

- Which face is fake?

- Unsupervised Learning:
  - **Data:** $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ samples (no labels!)
  - **Goal:** Learn underlying structure of data, e.g., $Q_X \approx P_X$
  - **Examples:** **Generative models**, clustering, dim. reduction, etc.

- GANs [Goodfellow *et al*'14]:
  - Sample-to-sample

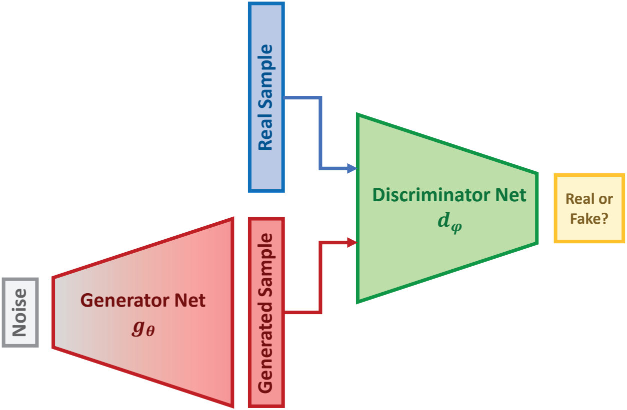
Generative Modeling - Preamble

- Which face is fake?

- **Unsupervised Learning:**
  - **Data:** $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ samples (no labels!)
  - **Goal:** Learn underlying structure of data, e.g., $Q_X \approx P_X$
  - **Examples:** **Generative models**, clustering, dim. reduction, etc.

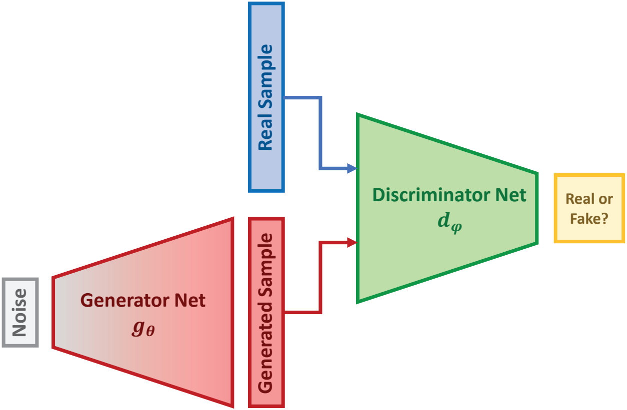
- **GANs [Goodfellow *et al*'14]:**
  - Sample-to-sample
  - Artwork, coloring, super-resolution, simulations, etc.

Designing Generative Adversarial Networks



Designing Generative Adversarial Networks

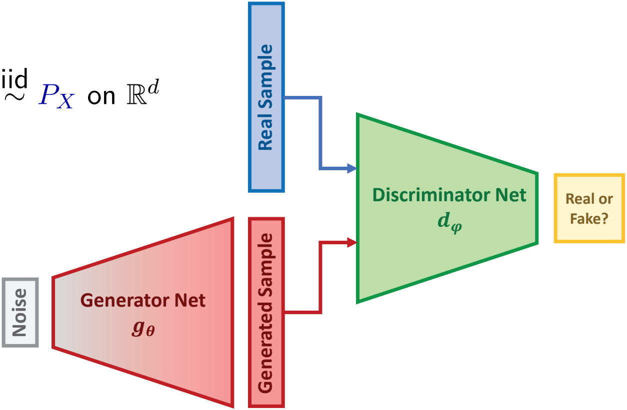
Resources:



Designing Generative Adversarial Networks

Resources:

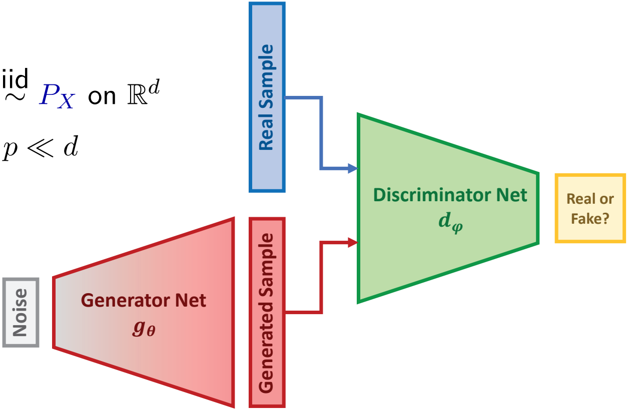
1. Data (real) $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ on $\mathbb{R}^d$



Designing Generative Adversarial Networks

Resources:

1. Data (real) $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ on $\mathbb{R}^d$
2. Noise $Z \sim \mathcal{N}(0, \mathrm{I}_p)$, $p \ll d$

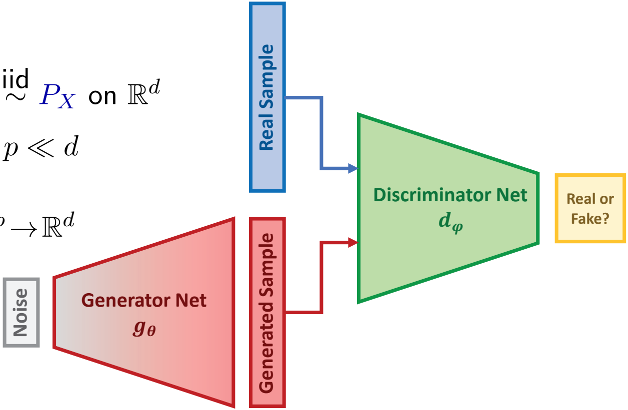


Designing Generative Adversarial Networks

Resources:

1. Data (real) $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ on $\mathbb{R}^d$
2. Noise $Z \sim \mathcal{N}(0, \mathrm{I}_p)$, $p \ll d$

Generator: DNN $g_\theta : \mathbb{R}^p \to \mathbb{R}^d$



Designing Generative Adversarial Networks

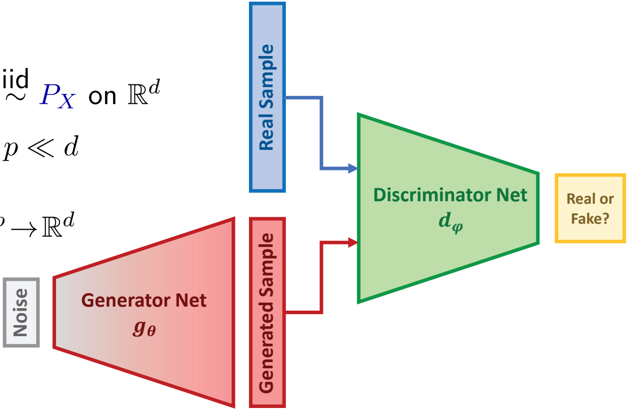
Resources:

1. Data (real) $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ on $\mathbb{R}^d$
2. Noise $Z \sim \mathcal{N}(0, \mathrm{I}_p)$, $p \ll d$

Generator: DNN $g_\theta : \mathbb{R}^p \to \mathbb{R}^d$

- Fake sample:

  $X_g = g_\theta(Z)$



Designing Generative Adversarial Networks

Resources:

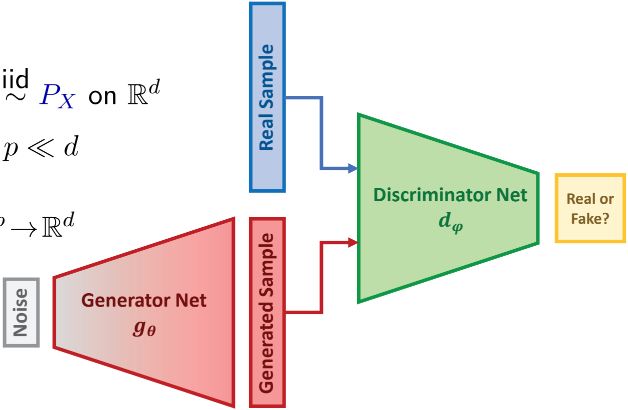
1. Data (real) $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ on $\mathbb{R}^d$
2. Noise $Z \sim \mathcal{N}(0, \mathrm{I}_p)$, $p \ll d$

Generator: DNN $g_\theta : \mathbb{R}^p \to \mathbb{R}^d$

- Fake sample:

  $X_g = g_\theta(Z)$

  s.t. $X_g \sim Q_{X_g}^{(\theta)} \approx P_X$



Designing Generative Adversarial Networks

Resources:

1. Data (real) $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ on $\mathbb{R}^d$
2. Noise $Z \sim \mathcal{N}(0, \mathrm{I}_p)$, $p \ll d$

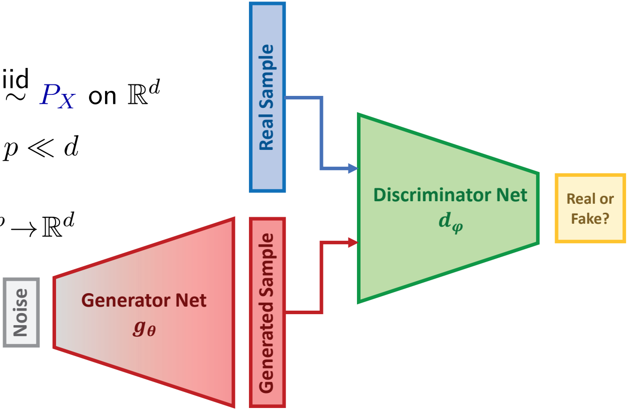
Generator: DNN $g_\theta : \mathbb{R}^p \to \mathbb{R}^d$

- Fake sample:

  $X_g = g_\theta(Z)$

  s.t. $X_g \sim Q_{X_g}^{(\theta)} \approx P_X$

Discriminator: DNN $d_\varphi : \mathbb{R}^d \to \mathbb{R}$



Designing Generative Adversarial Networks

Resources:

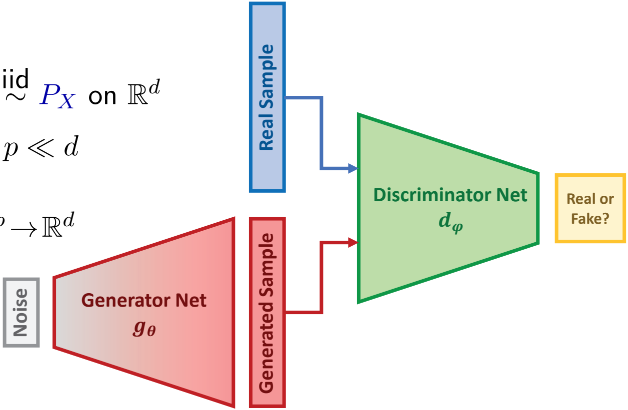
1. Data (real) $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ on $\mathbb{R}^d$
2. Noise $Z \sim \mathcal{N}(0, \mathrm{I}_p)$, $p \ll d$

Generator: DNN $g_\theta : \mathbb{R}^p \to \mathbb{R}^d$

- Fake sample:

  $X_g = g_\theta(Z)$

  s.t. $X_g \sim Q_{X_g}^{(\theta)} \approx P_X$



Discriminator: DNN $d_\varphi : \mathbb{R}^d \to \mathbb{R}$

- Discriminate: $d_\varphi(X)$ and $d_\varphi(X_g)$ are maximally different

Designing Generative Adversarial Networks

Resources:

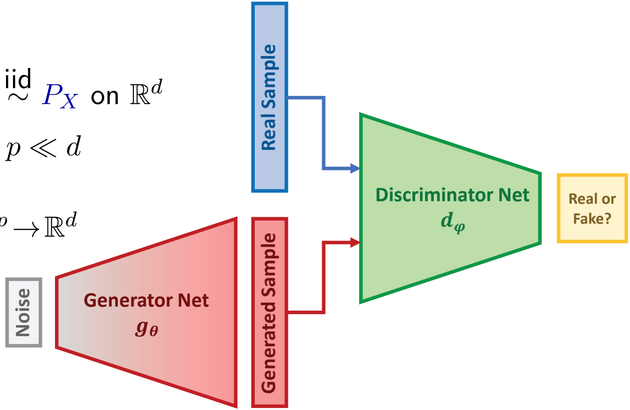
1. Data (real) $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ on $\mathbb{R}^d$
2. Noise $Z \sim \mathcal{N}(0, \mathrm{I}_p)$, $p \ll d$

Generator: DNN $g_\theta : \mathbb{R}^p \to \mathbb{R}^d$

- Fake sample:

  $X_g = g_\theta(Z)$

  s.t. $X_g \sim Q_{X_g}^{(\theta)} \approx P_X$



Discriminator: DNN $d_\varphi : \mathbb{R}^d \to \mathbb{R}$

- Discriminate: $d_\varphi(X)$ and $d_\varphi(X_g)$ are maximally different

$\Longrightarrow$ **Minimax Game:**

Designing Generative Adversarial Networks

Resources:

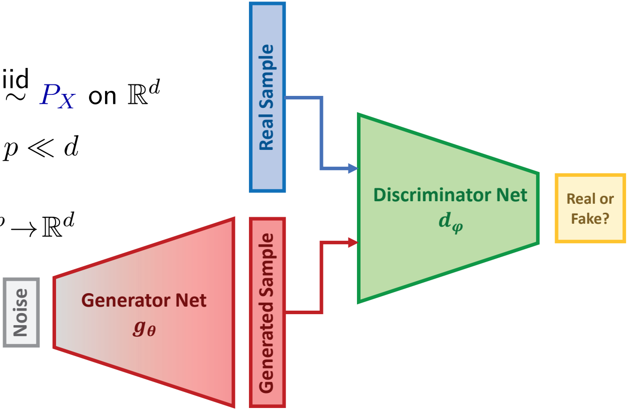
1. Data (real) $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ on $\mathbb{R}^d$
2. Noise $Z \sim \mathcal{N}(0, \mathrm{I}_p)$, $p \ll d$

Generator: DNN $g_\theta : \mathbb{R}^p \to \mathbb{R}^d$

- Fake sample:

  $X_g = g_\theta(Z)$

  s.t. $X_g \sim Q_{X_g}^{(\theta)} \approx P_X$



Discriminator: DNN $d_\varphi : \mathbb{R}^d \to \mathbb{R}$

- Discriminate: $d_\varphi(X)$ and $d_\varphi(X_g)$ are maximally different

$\Longrightarrow$ **Minimax Game:** $\mathbb{E} d_\varphi(X)$

Designing Generative Adversarial Networks

Resources:

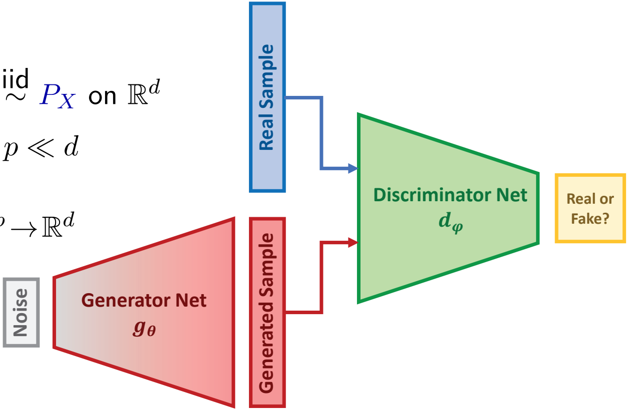
1. Data (real) $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ on $\mathbb{R}^d$
2. Noise $Z \sim \mathcal{N}(0, \mathrm{I}_p)$, $p \ll d$

Generator: DNN $g_\theta : \mathbb{R}^p \to \mathbb{R}^d$

- Fake sample:

  $X_g = g_\theta(Z)$

  s.t. $X_g \sim Q_{X_g}^{(\theta)} \approx P_X$



Discriminator: DNN $d_\varphi : \mathbb{R}^d \to \mathbb{R}$

- Discriminate: $d_\varphi(X)$ and $d_\varphi(X_g)$ are maximally different

$\Longrightarrow$ **Minimax Game:** $\mathbb{E}d_\varphi(X) - \mathbb{E}d_\varphi(g_\theta(Z))$

Designing Generative Adversarial Networks

Resources:

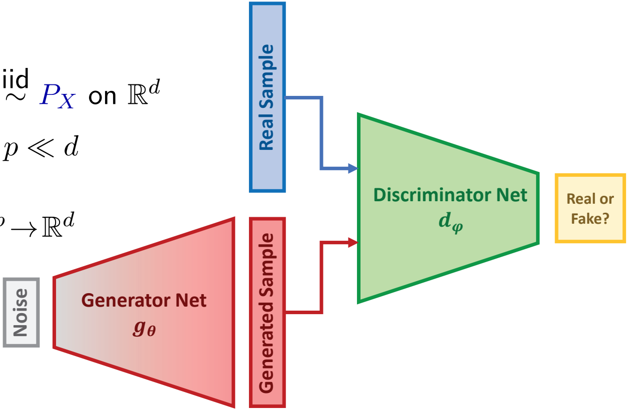
1. Data (real) $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ on $\mathbb{R}^d$
2. Noise $Z \sim \mathcal{N}(0, \mathrm{I}_p)$, $p \ll d$

Generator: DNN $g_\theta : \mathbb{R}^p \to \mathbb{R}^d$

- Fake sample:

  $X_g = g_\theta(Z)$

  s.t. $X_g \sim Q_{X_g}^{(\theta)} \approx P_X$



Discriminator: DNN $d_\varphi : \mathbb{R}^d \to \mathbb{R}$

- Discriminate: $d_\varphi(X)$ and $d_\varphi(X_g)$ are maximally different

$\Longrightarrow$ **Minimax Game:** $\sup_\varphi \mathbb{E} d_\varphi(X) - \mathbb{E} d_\varphi(g_\theta(Z))$

Designing Generative Adversarial Networks

Resources:

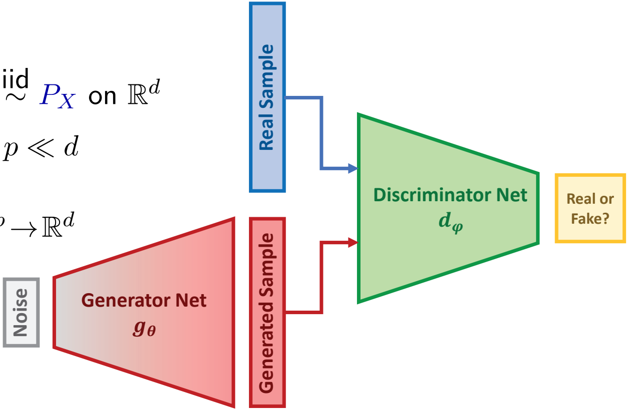
1. Data (real) $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ on $\mathbb{R}^d$
2. Noise $Z \sim \mathcal{N}(0, \mathrm{I}_p)$, $p \ll d$

Generator: DNN $g_\theta : \mathbb{R}^p \to \mathbb{R}^d$

- Fake sample:

  $X_g = g_\theta(Z)$

  s.t. $X_g \sim Q_{X_g}^{(\theta)} \approx P_X$



Discriminator: DNN $d_\varphi : \mathbb{R}^d \to \mathbb{R}$

- Discriminate: $d_\varphi(X)$ and $d_\varphi(X_g)$ are maximally different

$\Longrightarrow$ **Minimax Game:** $\inf_\theta \sup_\varphi \mathbb{E} d_\varphi(X) - \mathbb{E} d_\varphi(g_\theta(Z))$

Designing Generative Adversarial Networks

Resources:

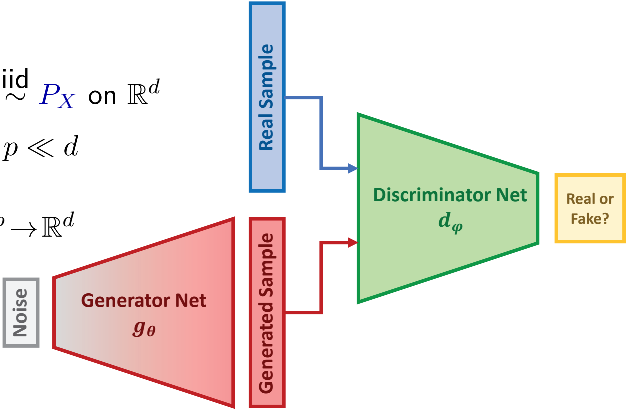
1. Data (real) $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} P_X$ on $\mathbb{R}^d$
2. Noise $Z \sim \mathcal{N}(0, \mathrm{I}_p)$, $p \ll d$

Generator: DNN $g_\theta : \mathbb{R}^p \to \mathbb{R}^d$

- Fake sample:

  $X_g = g_\theta(Z)$

  s.t. $X_g \sim Q_{X_g}^{(\theta)} \approx P_X$



Discriminator: DNN $d_\varphi : \mathbb{R}^d \to \mathbb{R}$

- Discriminate: $d_\varphi(X)$ and $d_\varphi(X_g)$ are maximally different

$\Longrightarrow$ **Minimax Game:** $\boxed{\inf_\theta \sup_\varphi \mathbb{E} d_\varphi(X) - \mathbb{E} d_\varphi(g_\theta(Z))}$

A Principled Approach for Generative Modeling

Statistical Divergences: Measure ‘distance’ between prob. distributions

A Principled Approach for Generative Modeling

Statistical Divergences: Measure 'distance' between prob. distributions

- $\mathcal{P}(\mathbb{R}^d) = \{\text{Set of probability distributions on } \mathbb{R}^d\}$

A Principled Approach for Generative Modeling

Statistical Divergences: Measure 'distance' between prob. distributions

- $\mathcal{P}(\mathbb{R}^d) = \{$Set of probability distributions on $\mathbb{R}^d\}$
- $\delta : \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d) \to [0, +\infty)$ is a divergence if

$$\delta(P, Q) = 0 \quad \Longleftrightarrow \quad P = Q$$

A Principled Approach for Generative Modeling

Statistical Divergences: Measure 'distance' between prob. distributions

- $\mathcal{P}(\mathbb{R}^d) = \{$Set of probability distributions on $\mathbb{R}^d\}$
- $\delta : \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d) \to [0, +\infty)$ is a divergence if

$$\delta(P, Q) = 0 \quad \iff \quad P = Q$$

- If symmetric & $\delta(P, Q) \leq \delta(P, R) + \delta(R, Q)$ then δ is a metric!

A Principled Approach for Generative Modeling

Statistical Divergences: Measure 'distance' between prob. distributions

- $\mathcal{P}(\mathbb{R}^d) = \{$Set of probability distributions on $\mathbb{R}^d\}$
- $\delta : \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d) \to [0, +\infty)$ is a divergence if

$$\delta(P, Q) = 0 \quad \iff \quad P = Q$$

- If symmetric & $\delta(P, Q) \leq \delta(P, R) + \delta(R, Q)$ then δ is a metric!

Back to Generative Modeling: Pick $\delta(\cdot, \cdot)$ and train generator to

$$\inf_{\theta} \delta\Big(P_X, Q_{X_d}^{(\theta)}\Big)$$

A Principled Approach for Generative Modeling

Statistical Divergences: Measure 'distance' between prob. distributions

- $\mathcal{P}(\mathbb{R}^d) = \{$Set of probability distributions on $\mathbb{R}^d\}$
- $\delta : \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d) \to [0, +\infty)$ is a divergence if

$$\delta(P, Q) = 0 \quad \iff \quad P = Q$$

- If symmetric & $\delta(P, Q) \leq \delta(P, R) + \delta(R, Q)$ then δ is a metric!

Back to Generative Modeling: Pick $\delta(\cdot, \cdot)$ and train generator to

$$\inf_\theta \delta\Big(P_X, Q_{X_d}^{(\theta)}\Big)$$

⊛ Both perspectives coincide if $\delta(\cdot, \cdot)$ is the 1-Wasserstein metric

A Principled Approach for Generative Modeling

Statistical Divergences: Measure 'distance' between prob. distributions

- $\mathcal{P}(\mathbb{R}^d) = \{\text{Set of probability distributions on } \mathbb{R}^d\}$
- $\delta : \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d) \to [0, +\infty)$ is a divergence if

$$\delta(P, Q) = 0 \quad \iff \quad P = Q$$

- If symmetric & $\delta(P, Q) \leq \delta(P, R) + \delta(R, Q)$ then δ is a metric!

Back to Generative Modeling: Pick $\delta(\cdot, \cdot)$ and train generator to

$$\inf_\theta \delta\left(P_X, Q_{X_d}^{(\theta)}\right)$$

- Both perspectives coincide if $\delta(\cdot, \cdot)$ is the 1-Wasserstein metric
- Wasserstein GAN achieves SOTA performance [Arjovsky *et al*'17]

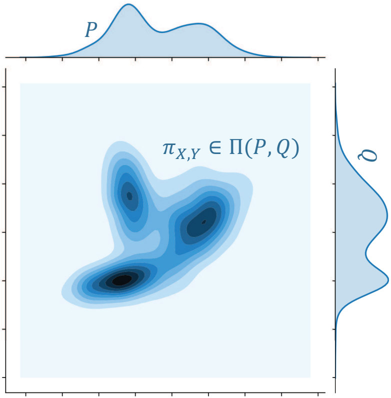
Optimal Transport: The 1-Wasserstein Metric

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

Optimal Transport: The 1-Wasserstein Metric

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

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



Optimal Transport: The 1-Wasserstein Metric

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

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

Optimal Transport: The 1-Wasserstein Metric

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

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

Definition (1-Wasserstein)

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

Optimal Transport: The 1-Wasserstein Metric

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

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

Definition (1-Wasserstein)

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

Comments:

Optimal Transport: The 1-Wasserstein Metric

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

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

Definition (1-Wasserstein)

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

Comments:

- **Operational Meaning:** Minimize work of transporting P to Q

Optimal Transport: The 1-Wasserstein Metric

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

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

Definition (1-Wasserstein)

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

Comments:

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

Optimal Transport: The 1-Wasserstein Metric

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

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

Definition (1-Wasserstein)

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

Comments:

- **Operational Meaning:** Minimize work of transporting P to Q
- **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)

The 1-Wasserstein Metric and GANs

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

The 1-Wasserstein Metric and GANs

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

Kantorovich-Rubinstein Duality: Equivalent representation

The 1-Wasserstein Metric and GANs

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

Kantorovich-Rubinstein Duality: Equivalent representation

$$\mathsf{W}_1(P,Q) = \sup_{\|f\|_{\mathsf{Lip}} \leq 1} \mathbb{E}_P f(X) - \mathbb{E}_Q f(Y)$$

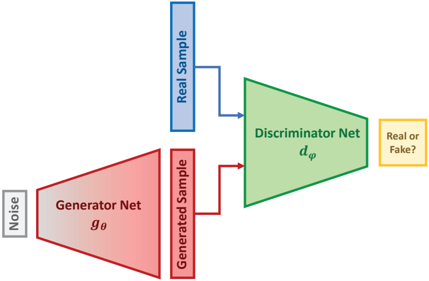
The 1-Wasserstein Metric and GANs

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

Kantorovich-Rubinstein Duality: Equivalent representation

$$\mathsf{W}_1(P,Q) = \sup_{\|f\|_{\mathsf{Lip}}\le 1} \mathbb{E}_P f(X) - \mathbb{E}_Q f(Y)$$

Back to GANs:



The 1-Wasserstein Metric and GANs

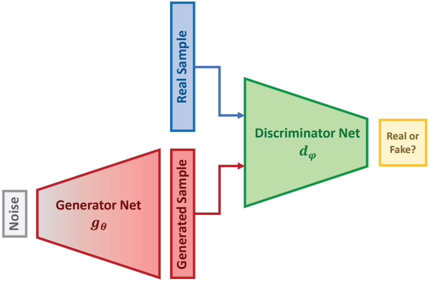
1-Wasserstein: $\mathsf{W}_1(P,Q) \triangleq \inf_{\pi_{X,Y}\in\Pi(P,Q)} \mathbb{E}_\pi\|X-Y\|$

Kantorovich-Rubinstein Duality: Equivalent representation

$$\mathsf{W}_1(\boldsymbol{P},Q) = \sup_{\|f\|_{\mathsf{Lip}}\leq 1} \mathbb{E}_{\boldsymbol{P}} f(\boldsymbol{X}) - \mathbb{E}_Q f(Y)$$

Back to GANs:

- $P = P_X$ (X (real) data sample)



The 1-Wasserstein Metric and GANs

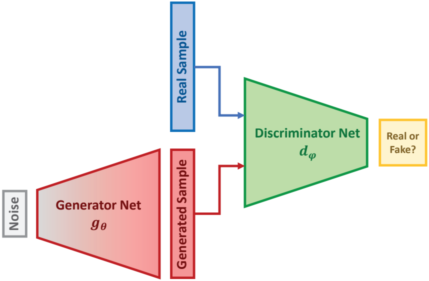
1-Wasserstein: $\mathsf{W}_1(P,Q) \triangleq \inf_{\pi_{X,Y}\in\Pi(P,Q)} \mathbb{E}_\pi \|X-Y\|$

Kantorovich-Rubinstein Duality: Equivalent representation

$$\mathsf{W}_1(\boldsymbol{P},\boldsymbol{Q}) = \sup_{\|f\|_{\mathsf{Lip}}\leq 1} \mathbb{E}_{\boldsymbol{P}} f(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{Q}} f(\boldsymbol{Y})$$

Back to GANs:

- $P = P_X$ (X (real) data sample)
- $Q = Q_{X_g}^{(\theta)}$ ($Y = g_\theta(Z)$ gen. sample)



The 1-Wasserstein Metric and GANs

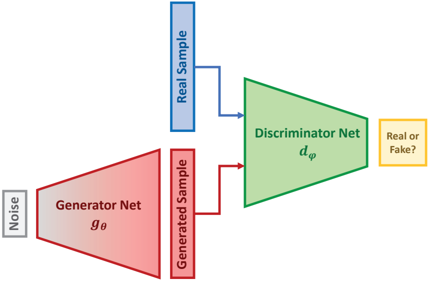
1-Wasserstein: $\mathsf{W}_1(P,Q) \triangleq \inf_{\pi_{X,Y}\in\Pi(P,Q)} \mathbb{E}_\pi \|X-Y\|$

Kantorovich-Rubinstein Duality: Equivalent representation

$$\mathsf{W}_1(\boldsymbol{P},\boldsymbol{Q}) = \sup_{\|f\|_{\mathsf{Lip}}\leq 1} \mathbb{E}_{\boldsymbol{P}} \boldsymbol{f}(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{Q}} \boldsymbol{f}(\boldsymbol{Y})$$

Back to GANs:

- $P = P_X$ (X (real) data sample)
- $Q = Q_{X_g}^{(\theta)}$ ($Y = g_\theta(Z)$ gen. sample)
- $f = d_\varphi$ (w/ 1-Lip constraint)



The 1-Wasserstein Metric and GANs

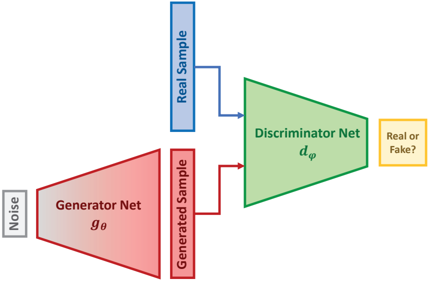
1-Wasserstein: $\mathsf{W}_1(P,Q) \triangleq \inf_{\pi_{X,Y}\in\Pi(P,Q)} \mathbb{E}_\pi \|X-Y\|$

Kantorovich-Rubinstein Duality: Equivalent representation

$$\mathsf{W}_1(\boldsymbol{P},\boldsymbol{Q}) = \sup_{\|f\|_{\mathsf{Lip}}\leq 1} \mathbb{E}_{\boldsymbol{P}} \boldsymbol{f}(\boldsymbol{X}) - \mathbb{E}_{\boldsymbol{Q}} \boldsymbol{f}(\boldsymbol{Y})$$

Back to GANs:

- $P = P_X$ (X (real) data sample)
- $Q = Q_{X_g}^{(\theta)}$ ($Y = g_\theta(Z)$ gen. sample)
- $f = d_\varphi$ (w/ 1-Lip constraint)



$\Longrightarrow$ **Frameworks Coincide:**

$$\inf_\theta \mathsf{W}_1\left(P_X, Q_{X_d}^{(\theta)}\right) \cong \inf_\theta \sup_{\varphi:\ \|d_\varphi\|_{\mathsf{Lip}}\leq 1} \mathbb{E} d_\varphi(X) - \mathbb{E} d_\varphi(g_\theta(Z))$$

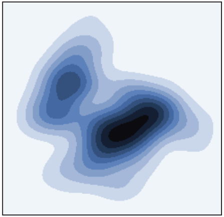
So Problem Solved...? Not in High-Dimensions

Goal: Find gen. g_θ to minimize $\mathsf{W}_1\left(P_X, Q_{X_d}^{(\theta)}\right)$, $X_d^{(\theta)} \triangleq g_\theta(Z) \sim Q_{X_d}^{(\theta)}$.

So Problem Solved...? Not in High-Dimensions

Goal: Find gen. g_θ to minimize $\mathsf{W}_1\Big(P_X, Q^{(\theta)}_{X_d}\Big)$, $X^{(\theta)}_d \triangleq g_\theta(Z) \sim Q^{(\theta)}_{X_d}$.

Empirical Approximation: In reality we don't have P_X but data

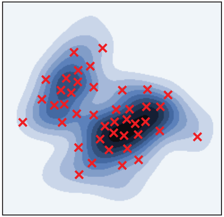


So Problem Solved...? Not in High-Dimensions

Goal: Find gen. g_θ to minimize $\mathsf{W}_1\Big(P_X, Q^{(\theta)}_{X_d}\Big)$, $X^{(\theta)}_d \triangleq g_\theta(Z) \sim Q^{(\theta)}_{X_d}$.

Empirical Approximation: In reality we don't have P_X but data

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

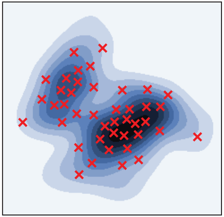


So Problem Solved...? Not in High-Dimensions

Goal: Find gen. g_θ to minimize $\mathsf{W}_1\Big(P_X, Q^{(\theta)}_{X_d}\Big)$, $X_d^{(\theta)} \triangleq g_\theta(Z) \sim Q^{(\theta)}_{X_d}$.

Empirical Approximation: In reality we don't have P_X but data

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

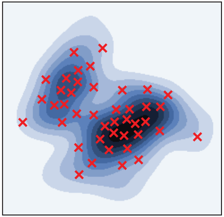


So Problem Solved...? Not in High-Dimensions

Goal: Find gen. g_θ to minimize $\mathsf{W}_1\left(P_X, Q^{(\theta)}_{X_d}\right)$, $X^{(\theta)}_d \triangleq g_\theta(Z) \sim Q^{(\theta)}_{X_d}$.

Empirical Approximation: In reality we don't have P_X but data

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



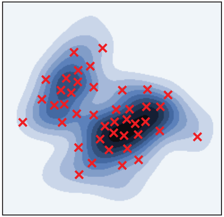
$\implies$ Inherently we work with $\mathsf{W}_1\left(\hat{P}_n, Q^{(\theta)}_{X_d}\right)$

So Problem Solved...? Not in High-Dimensions

Goal: Find gen. g_θ to minimize $\mathsf{W}_1\left(P_X, Q^{(\theta)}_{X_d}\right)$, $X_d^{(\theta)} \triangleq g_\theta(Z) \sim Q^{(\theta)}_{X_d}$.

Empirical Approximation: In reality we don't have P_X but data

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



$\Longrightarrow$ Inherently we work with $\mathsf{W}_1\left(\hat{P}_n, Q^{(\theta)}_{X_d}\right)$

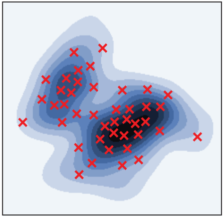
$$\left[\mathsf{W}_1\left(\hat{P}_n, Q^{(\theta)}_{X_d}\right) \approx \mathsf{W}_1\left(P_X, Q^{(\theta)}_{X_d}\right) \text{ hopefully...}\right]$$

So Problem Solved...? Not in High-Dimensions

Goal: Find gen. g_θ to minimize $\mathsf{W}_1\left(P_X, Q^{(\theta)}_{X_d}\right)$, $X^{(\theta)}_d \triangleq g_\theta(Z) \sim Q^{(\theta)}_{X_d}$.

Empirical Approximation: In reality we don't have P_X but data

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



$\Longrightarrow$ Inherently we work with $\mathsf{W}_1\left(\hat{P}_n, Q^{(\theta)}_{X_d}\right)$

$$\left[\mathsf{W}_1\left(\hat{P}_n, Q^{(\theta)}_{X_d}\right) \approx \mathsf{W}_1\left(P_X, Q^{(\theta)}_{X_d}\right) \text{ hopefully...}\right]$$

Theorem (Dudley'69)

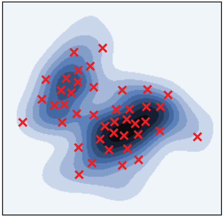
For all $d \geq 3$, $\mathbb{E}\left|W_1(\hat{P}_n, Q) - W_1(P, Q)\right| \asymp n^{-\frac{1}{d}}$

So Problem Solved...? Not in High-Dimensions

Goal: Find gen. g_θ to minimize $\mathsf{W}_1\left(P_X, Q^{(\theta)}_{X_d}\right)$, $X^{(\theta)}_d \triangleq g_\theta(Z) \sim Q^{(\theta)}_{X_d}$.

Empirical Approximation: In reality we don't have P_X but data

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



$\Longrightarrow$ Inherently we work with $\mathsf{W}_1\left(\hat{P}_n, Q^{(\theta)}_{X_d}\right)$

$$\left[\mathsf{W}_1\left(\hat{P}_n, Q^{(\theta)}_{X_d}\right) \approx \mathsf{W}_1\left(P_X, Q^{(\theta)}_{X_d}\right) \text{ hopefully...}\right]$$

Theorem (Dudley'69)

For all $d \geq 3$, $\mathbb{E}\left|W_1(\hat{P}_n, Q) - W_1(P, Q)\right| \asymp n^{-\frac{1}{d}}$

Curse of Dimensionality

Entropic Optimal Transport: For $\epsilon > 0$ and cost $c(x, y)$:

$$\mathsf{S}_c^{(\epsilon)}(P, Q) \triangleq \inf_{\pi_{X,Y} \in \Pi(P,Q)} \mathbb{E}_\pi c(X, Y) + \epsilon I_\pi(X; Y)$$

where $I_\pi(X; Y) \triangleq \mathsf{D}_{\mathsf{KL}}(\pi_{X,Y} \| P \otimes Q)$ is the mutual information

Curse of Dimensionality - Current State of Affairs

Entropic Optimal Transport: For $\epsilon > 0$ and cost $c(x,y)$:

$$\mathsf{S}_c^{(\epsilon)}(P,Q) \triangleq \inf_{\pi_{X,Y} \in \Pi(P,Q)} \mathbb{E}_\pi c(X,Y) + \epsilon I_\pi(X;Y)$$

where $I_\pi(X;Y) \triangleq \mathsf{D}_{\mathsf{KL}}(\pi_{X,Y} \| P \otimes Q)$ is the mutual information

The Good:

Curse of Dimensionality - Current State of Affairs

Entropic Optimal Transport: For $\epsilon > 0$ and cost $c(x,y)$:

$$\mathsf{S}_c^{(\epsilon)}(P,Q) \triangleq \inf_{\pi_{X,Y}\in\Pi(P,Q)} \mathbb{E}_\pi c(X,Y) + \epsilon I_\pi(X;Y)$$

where $I_\pi(X;Y) \triangleq \mathsf{D}_{\mathsf{KL}}(\pi_{X,Y}\|P\otimes Q)$ is the mutual information

The Good:

- Regularizer favors weakly-dependent coupling

Curse of Dimensionality - Current State of Affairs

Entropic Optimal Transport: For $\epsilon > 0$ and cost $c(x, y)$:

$$\mathsf{S}_c^{(\epsilon)}(P, Q) \triangleq \inf_{\pi_{X,Y} \in \Pi(P,Q)} \mathbb{E}_\pi c(X, Y) + \epsilon I_\pi(X; Y)$$

where $I_\pi(X; Y) \triangleq \mathsf{D}_{\mathsf{KL}}(\pi_{X,Y} \| P \otimes Q)$ is the mutual information

The Good:

- Regularizer favors weakly-dependent coupling
- $\mathsf{S}_c^{(\epsilon)}(P, Q)$ is strongly convex optimization problem

Curse of Dimensionality - Current State of Affairs

Entropic Optimal Transport: For $\epsilon > 0$ and cost $c(x,y)$:

$$\mathsf{S}_c^{(\epsilon)}(P,Q) \triangleq \inf_{\pi_{X,Y}\in\Pi(P,Q)} \mathbb{E}_\pi c(X,Y) + \epsilon I_\pi(X;Y)$$

where $I_\pi(X;Y) \triangleq \mathsf{D}_{\mathsf{KL}}(\pi_{X,Y}\|P\otimes Q)$ is the mutual information

The Good:

- Regularizer favors weakly-dependent coupling
- $\mathsf{S}_c^{(\epsilon)}(P,Q)$ is strongly convex optimization problem
- Two-sample statistical efficiency (for certain costs):

Curse of Dimensionality - Current State of Affairs

Entropic Optimal Transport: For $\epsilon > 0$ and cost $c(x,y)$:

$$\mathsf{S}_c^{(\epsilon)}(P,Q) \triangleq \inf_{\pi_{X,Y}\in\Pi(P,Q)} \mathbb{E}_\pi c(X,Y) + \epsilon I_\pi(X;Y)$$

where $I_\pi(X;Y) \triangleq \mathsf{D}_{\mathsf{KL}}(\pi_{X,Y}\|P\otimes Q)$ is the mutual information

The Good:

- Regularizer favors weakly-dependent coupling
- $\mathsf{S}_c^{(\epsilon)}(P,Q)$ is strongly convex optimization problem
- Two-sample statistical efficiency (for certain costs):

Theorem (Genevay et al'19)

For $\mathcal{C}^\infty$ and L-Lipschitz cost c, and any $d \geq 1$, $\epsilon > 0$:

$$\mathbb{E}\left|S_c^{(\epsilon)}(\hat{P}_n,\hat{Q}_n) - S_c^{(\epsilon)}(P,Q)\right| \lesssim e^{\frac{c}{\epsilon}}\left(1+\frac{1}{\epsilon^{\lfloor d/2\rfloor}}\right)n^{-\frac{1}{2}}$$

Curse of Dimensionality - Current State of Affairs

Entropic Optimal Transport: For $\epsilon > 0$ and cost $c(x,y)$:

$$\mathsf{S}_c^{(\epsilon)}(P,Q) \triangleq \inf_{\pi_{X,Y} \in \Pi(P,Q)} \mathbb{E}_\pi c(X,Y) + \epsilon I_\pi(X;Y)$$

where $I_\pi(X;Y) \triangleq \mathsf{D}_{\mathsf{KL}}(\pi_{X,Y} \| P \otimes Q)$ is the mutual information

The Bad (Specializing to Distance Cost):

Curse of Dimensionality - Current State of Affairs

Entropic Optimal Transport: For $\epsilon > 0$ and cost $c(x, y)$:

$$\mathsf{S}_c^{(\epsilon)}(P, Q) \triangleq \inf_{\pi_{X,Y} \in \Pi(P,Q)} \mathbb{E}_\pi c(X, Y) + \epsilon I_\pi(X; Y)$$

where $I_\pi(X; Y) \triangleq \mathsf{D}_{\mathsf{KL}}(\pi_{X,Y} \| P \otimes Q)$ is the mutual information

The Bad (Specializing to Distance Cost):

- $c(x, y) = \|x - y\|$ does not fall under the theorem's framework

Curse of Dimensionality - Current State of Affairs

Entropic Optimal Transport: For $\epsilon > 0$ and cost $c(x,y)$:

$$\mathsf{S}_c^{(\epsilon)}(P,Q) \triangleq \inf_{\pi_{X,Y} \in \Pi(P,Q)} \mathbb{E}_\pi c(X,Y) + \epsilon I_\pi(X;Y)$$

where $I_\pi(X;Y) \triangleq \mathsf{D}_{\mathsf{KL}}(\pi_{X,Y} \| P \otimes Q)$ is the mutual information

The Bad (Specializing to Distance Cost):

- $c(x,y) = \|x-y\|$ does not fall under the theorem's framework
- $\mathsf{S}_1^{(\epsilon)}$ is not a metric on $\mathcal{P}_1(\mathbb{R}^d)$ for any $\epsilon > 0$ (nor Sinkhorn loss)

Curse of Dimensionality - Current State of Affairs

Entropic Optimal Transport: For $\epsilon > 0$ and cost $c(x, y)$:

$$\mathsf{S}_c^{(\epsilon)}(P, Q) \triangleq \inf_{\pi_{X,Y} \in \Pi(P,Q)} \mathbb{E}_\pi c(X, Y) + \epsilon I_\pi(X; Y)$$

where $I_\pi(X; Y) \triangleq \mathsf{D}_{\mathsf{KL}}(\pi_{X,Y} \| P \otimes Q)$ is the mutual information

The Bad (Specializing to Distance Cost):

- $c(x, y) = \|x - y\|$ does not fall under the theorem's framework
- $\mathsf{S}_1^{(\epsilon)}$ is not a metric on $\mathcal{P}_1(\mathbb{R}^d)$ for any $\epsilon > 0$ (nor Sinkhorn loss)
- Dual form includes two potentials and regularizer:

$$S_1^{(\epsilon)}(P, Q) = \sup_{u,v \in \mathcal{C}(\mathbb{R}^d)} \mathbb{E}_{P \otimes Q}\left[u(X) + v(Y) - \epsilon e^{\frac{u(X)+v(Y)-\|X-Y\|}{\epsilon}}\right] + \epsilon$$

Curse of Dimensionality - Current State of Affairs

Entropic Optimal Transport: For $\epsilon > 0$ and cost $c(x,y)$:

$$\mathsf{S}_c^{(\epsilon)}(P,Q) \triangleq \inf_{\pi_{X,Y} \in \Pi(P,Q)} \mathbb{E}_\pi c(X,Y) + \epsilon I_\pi(X;Y)$$

where $I_\pi(X;Y) \triangleq \mathsf{D}_{\mathsf{KL}}(\pi_{X,Y} \| P \otimes Q)$ is the mutual information

The Bad (Specializing to Distance Cost):

- $c(x,y) = \|x-y\|$ does not fall under the theorem's framework
- $\mathsf{S}_1^{(\epsilon)}$ is <u>not</u> a metric on $\mathcal{P}_1(\mathbb{R}^d)$ for any $\epsilon > 0$ (nor Sinkhorn loss)
- Dual form includes two potentials and regularizer:

$$S_1^{(\epsilon)}(P,Q) = \sup_{u,v \in \mathcal{C}(\mathbb{R}^d)} \mathbb{E}_{P \otimes Q}\left[u(X) + v(Y) - \epsilon e^{\frac{u(X)+v(Y)-\|X-Y\|}{\epsilon}}\right] + \epsilon$$

$\Longrightarrow$ No direct correspondence to minimax GAN formulation

Gaussian-Smoothed Optimal Transport

Definition (ZG-Greenewald'19)

For $\sigma \geq 0$, the Gaussian-smoothed OT (GOT) between P and Q is

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

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

Gaussian-Smoothed Optimal Transport

Definition (ZG-Greenewald'19)

For $\sigma \geq 0$, the Gaussian-smoothed OT (GOT) between P and Q is

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

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 $Z_1, Z_2 \sim \mathcal{N}_\sigma$

Gaussian-Smoothed Optimal Transport

Definition (ZG-Greenewald'19)

For $\sigma \geq 0$, the Gaussian-smoothed OT (GOT) between P and Q is

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

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 $Z_1, Z_2 \sim \mathcal{N}_\sigma$

$$X \perp Z_1 \quad \Longrightarrow \quad X + Z_1 \sim P * \mathcal{N}_\sigma$$

$$Y \perp Z_2 \quad \Longrightarrow \quad Y + Z_2 \sim Q * \mathcal{N}_\sigma$$

Gaussian-Smoothed Optimal Transport

Definition (ZG-Greenewald'19)

For $\sigma \geq 0$, the Gaussian-smoothed OT (GOT) between P and Q is

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

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 $Z_1, Z_2 \sim \mathcal{N}_\sigma$

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

$\Longrightarrow$ W_1 distance between smoothed distributions

Gaussian-Smoothed Optimal Transport

Definition (ZG-Greenewald'19)

For $\sigma \geq 0$, the Gaussian-smoothed OT (GOT) between P and Q is

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

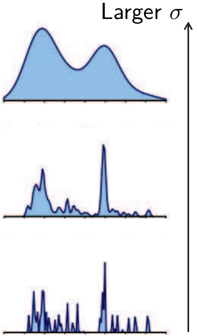
where $\mathcal{N}_\sigma \triangleq \mathcal{N}(0, \sigma^2 \mathrm{I}_d)$ is a d-dimensional isotropic Gaussian.

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

$$X \perp Z_1 \quad \Longrightarrow \quad X + Z_1 \sim P * \mathcal{N}_\sigma$$

$$Y \perp Z_2 \quad \Longrightarrow \quad Y + Z_2 \sim Q * \mathcal{N}_\sigma$$

$\Longrightarrow$ W_1 distance between smoothed distributions



Gaussian-Smoothed Optimal Transport

Definition (ZG-Greenewald'19)

For $\sigma \geq 0$, the Gaussian-smoothed OT (GOT) between P and Q is

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

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 $Z_1, Z_2 \sim \mathcal{N}_\sigma$

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

$\Longrightarrow$ W_1 distance between smoothed distributions

GAN-Compatibility: GOT is W_1 but between convolved distributions

Gaussian-Smoothed Optimal Transport

Definition (ZG-Greenewald'19)

For $\sigma \geq 0$, the Gaussian-smoothed OT (GOT) between P and Q is

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

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 $Z_1, Z_2 \sim \mathcal{N}_\sigma$

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

$\Longrightarrow$ W_1 distance between smoothed distributions

GAN-Compatibility: GOT is W_1 but between convolved distributions

- **KR Duality:** $\mathsf{W}_1^{(\sigma)}(P,Q) = \sup_{\|f\|_{\mathsf{Lip}} \leq 1} \mathbb{E} f(X+Z) - \mathbb{E} f(Y+Z)$

Gaussian-Smoothed OT - Metric Structure

High Level: GOT inherits the metric structure of 1-Wasserstein

Gaussian-Smoothed OT - Metric Structure

High Level: GOT 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.).*

Gaussian-Smoothed OT - Metric Structure

High Level: GOT 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 Proof: Use Characteristic functions $\phi_P(t) \triangleq \mathbb{E}_P\left[e^{itX}\right]$ and:

Gaussian-Smoothed OT - Metric Structure

High Level: GOT 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 Proof: Use Characteristic functions $\phi_P(t) \triangleq \mathbb{E}_P\left[e^{itX}\right]$ and:

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

Gaussian-Smoothed OT - Metric Structure

High Level: GOT 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 Proof: 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'19)

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

Gaussian-Smoothed OT - Metric Structure

High Level: GOT 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 Proof: 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'19)

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

⊛ **GOT induces exact same topology as classic Wasserstein**

Gaussian-Smoothed OT - Function of Noise Std

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

Gaussian-Smoothed OT - Function of Noise Std

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:

Gaussian-Smoothed OT - Function of Noise Std

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* $\sigma \in [0,+\infty)$

Gaussian-Smoothed OT - Function of Noise Std

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. $\lim_{\sigma\to 0} \mathsf{W}_1^{(\sigma)}(P,Q) = \mathsf{W}_1(P,Q)$

Gaussian-Smoothed OT - Function of Noise Std

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* $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$

Gaussian-Smoothed OT - Function of Noise Std

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* $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$

Key Idea for Proof: Use dual form to relate $\mathsf{W}_1^{(\sigma)}$ and W_1 as:

Gaussian-Smoothed OT - Function of Noise Std

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* $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$

Key Idea for Proof: Use dual form to relate $\mathsf{W}_1^{(\sigma)}$ and W_1 as:

Lemma (ZG-Greenewald'19)

Fix $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$, and $0 \le \sigma_1 < \sigma_2 < +\infty$. We have

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

Gaussian-Smoothed OT - Function of Noise Std

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* $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$

Third Item: Intuitively should decay to 0?

Gaussian-Smoothed OT - Function of Noise Std

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* $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$

Third Item: Intuitively should decay to 0? Let $P=\delta_x,\ Q=\delta_y,\ x\neq y$

Gaussian-Smoothed OT - Function of Noise Std

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* $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$

Third Item: Intuitively should decay to 0? Let $P=\delta_x,\ Q=\delta_y,\ x\neq y$

$$\mathsf{W}_1^{(\sigma)}(P,Q) = \mathsf{W}_1\big(\mathcal{N}(x,\sigma^2\mathrm{I}_d),\mathcal{N}(y,\sigma^2\mathrm{I}_d)\big)$$

Gaussian-Smoothed OT - Function of Noise Std

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* $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$

Third Item: Intuitively should decay to 0? Let $P=\delta_x,\ Q=\delta_y,\ x\neq y$

$$\begin{aligned}\mathsf{W}_1^{(\sigma)}(P,Q) &= \mathsf{W}_1\big(\mathcal{N}(x,\sigma^2\mathrm{I}_d),\mathcal{N}(y,\sigma^2\mathrm{I}_d)\big)\\ &\geq \Big\|\mathbb{E}_{\mathcal{N}(x,\sigma^2\mathrm{I}_d)}X - \mathbb{E}_{\mathcal{N}(y,\sigma^2\mathrm{I}_d)}Y\Big\|\end{aligned}$$

Gaussian-Smoothed OT - Function of Noise Std

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* $P,Q \in \mathcal{P}_1(\mathbb{R}^d)$

Third Item: Intuitively should decay to 0? Let $P=\delta_x,\ Q=\delta_y,\ x\neq y$

$$\begin{aligned}\mathsf{W}_1^{(\sigma)}(P,Q) &= \mathsf{W}_1\big(\mathcal{N}(x,\sigma^2\mathrm{I}_d),\mathcal{N}(y,\sigma^2\mathrm{I}_d)\big)\\ &\geq \Big\|\mathbb{E}_{\mathcal{N}(x,\sigma^2\mathrm{I}_d)}X - \mathbb{E}_{\mathcal{N}(y,\sigma^2\mathrm{I}_d)}Y\Big\|\\ &= \|x-y\|\end{aligned}$$

Gaussian-Smoothed OT - Function of Noise Std

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* $\sigma_k \searrow \sigma \geq 0$. *Let:*

Gaussian-Smoothed OT - Function of Noise Std

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 $\sigma_k \searrow \sigma \geq 0$. Let:

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

Gaussian-Smoothed OT - Function of Noise Std

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* $\sigma_k \searrow \sigma \geq 0$. *Let:*

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

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

Gaussian-Smoothed OT - Function of Noise Std

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* $\sigma_k \searrow \sigma \geq 0$. *Let:*

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

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

- $\pi \in \Pi(P * \mathcal{N}_\sigma, Q * \mathcal{N}_\sigma)$

Gaussian-Smoothed OT - Function of Noise Std

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* $\sigma_k \searrow \sigma \geq 0$. *Let:*

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

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

- $\pi \in \Pi(P * \mathcal{N}_\sigma, Q * \mathcal{N}_\sigma)$
- π *is optimal for* $\mathsf{W}_1^{(\sigma)}(\mu,\nu)$

Gaussian-Smoothed OT - Function of Noise Std

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 $\sigma_k \searrow \sigma \geq 0$. Let:

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

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

- $\pi \in \Pi(P * \mathcal{N}_\sigma, Q * \mathcal{N}_\sigma)$
- π *is optimal for* $\mathsf{W}_1^{(\sigma)}(\mu,\nu)$

Comments:

Gaussian-Smoothed OT - Function of Noise Std

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* $\sigma_k \searrow \sigma \geq 0$. *Let:*

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

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

- $\pi \in \Pi(P * \mathcal{N}_\sigma, Q * \mathcal{N}_\sigma)$
- π *is optimal for* $\mathsf{W}_1^{(\sigma)}(\mu,\nu)$

Comments:

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

Gaussian-Smoothed OT - Function of Noise Std

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* $\sigma_k \searrow \sigma \geq 0$. *Let:*

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

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

- $\pi \in \Pi(P * \mathcal{N}_\sigma, Q * \mathcal{N}_\sigma)$
- π *is optimal for* $\mathsf{W}_1^{(\sigma)}(\mu,\nu)$

Comments:

- **In words:** Not only opt. values converge, but also optimizers
- $\boldsymbol{\sigma = 0}$**:** Gaussian-smoothed opt. plans converge to W_1 opt. coupling.

Gaussian-Smoothed OT - Function of Noise Std

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* $\sigma_k \searrow \sigma \geq 0$. *Let:*

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

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

- $\pi \in \Pi(P * \mathcal{N}_\sigma, Q * \mathcal{N}_\sigma)$
- π *is optimal for* $\mathsf{W}_1^{(\sigma)}(\mu,\nu)$

Comments:

- **In words:** Not only opt. values converge, but also optimizers
- $\boldsymbol{\sigma = 0}$**:** Gaussian-smoothed opt. plans converge to W_1 opt. coupling.
- **Proof Idea:** Γ-convergence (CoV) & Tightness of $\Pi(\mu,\nu)$ (Topology)

Gaussian-Smoothed OT - Statistical Efficiency

High Level: Fast convergence of empirical approx. under $\mathsf{W}_1^{(\sigma)}$

Gaussian-Smoothed OT - Statistical Efficiency

High Level: Fast convergence of empirical approx. under $\mathsf{W}_1^{(\sigma)}$

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

For any $d \geq 1$ and $\sigma > 0$, we have: $\mathbb{E}\mathsf{W}_1^{(\sigma)}(\hat{P}_n, P) \lesssim \left(\frac{c}{\sigma}\right)^{\frac{d}{2}} n^{-\frac{1}{2}}$

Gaussian-Smoothed OT - Statistical Efficiency

High Level: Fast convergence of empirical approx. under $\mathsf{W}_1^{(\sigma)}$

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

For any $d \geq 1$ and $\sigma > 0$, we have: $\mathbb{E}\mathsf{W}_1^{(\sigma)}(\hat{P}_n, P) \lesssim \left(\frac{c}{\sigma}\right)^{\frac{d}{2}} n^{-\frac{1}{2}}$

Comments:

Gaussian-Smoothed OT - Statistical Efficiency

High Level: Fast convergence of empirical approx. under $\mathsf{W}_1^{(\sigma)}$

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

For any $d \geq 1$ and $\sigma > 0$, we have: $\mathbb{E}\mathsf{W}_1^{(\sigma)}(\hat{P}_n, P) \lesssim \left(\frac{c}{\sigma}\right)^{\frac{d}{2}} n^{-\frac{1}{2}}$

Comments:

- True for any subgauassian P (entropic OT assumes compact $\mathrm{supp}(P)$)

Gaussian-Smoothed OT - Statistical Efficiency

High Level: Fast convergence of empirical approx. under $\mathsf{W}_1^{(\sigma)}$

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

For any $d \geq 1$ and $\sigma > 0$, we have: $\mathbb{E}\mathsf{W}_1^{(\sigma)}(\hat{P}_n, P) \lesssim \left(\frac{c}{\sigma}\right)^{\frac{d}{2}} n^{-\frac{1}{2}}$

Comments:

- True for any subgauassian P (entropic OT assumes compact supp(P))
- Generalizes to subgauassian noise distributions with monotone density

Gaussian-Smoothed OT - Statistical Efficiency

High Level: Fast convergence of empirical approx. under $\mathsf{W}_1^{(\sigma)}$

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

For any $d \geq 1$ and $\sigma > 0$, we have: $\mathbb{E}\mathsf{W}_1^{(\sigma)}(\hat{P}_n, P) \lesssim \left(\frac{c}{\sigma}\right)^{\frac{d}{2}} n^{-\frac{1}{2}}$

Comments:

- True for any subgauassian P (entropic OT assumes compact supp(P))
- Generalizes to subgauassian noise distributions with monotone density
- Implies fast convergence of the other empirical approx. setups:

Gaussian-Smoothed OT - Statistical Efficiency

High Level: Fast convergence of empirical approx. under $\mathsf{W}_1^{(\sigma)}$

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

For any $d \geq 1$ and $\sigma > 0$, we have: $\mathbb{E}\mathsf{W}_1^{(\sigma)}(\hat{P}_n, P) \lesssim \left(\frac{c}{\sigma}\right)^{\frac{d}{2}} n^{-\frac{1}{2}}$

Comments:

- True for any subgauassian P (entropic OT assumes compact $\mathrm{supp}(P)$)
- Generalizes to subgauassian noise distributions with monotone density
- Implies fast convergence of the other empirical approx. setups:
  - **One-Sample:** $\mathbb{E}\left|\mathsf{W}_1^{(\sigma)}(\hat{P}_n, Q) - \mathsf{W}_1^{(\sigma)}(P, Q)\right| \in O\left(n^{-\frac{1}{2}}\right)$ (GANs)

Gaussian-Smoothed OT - Statistical Efficiency

High Level: Fast convergence of empirical approx. under $\mathsf{W}_1^{(\sigma)}$

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

For any $d \geq 1$ and $\sigma > 0$, we have: $\mathbb{E}\mathsf{W}_1^{(\sigma)}(\hat{P}_n, P) \lesssim \left(\frac{c}{\sigma}\right)^{\frac{d}{2}} n^{-\frac{1}{2}}$

Comments:

- True for any subgauassian P (entropic OT assumes compact supp(P))
- Generalizes to subgauassian noise distributions with monotone density
- Implies fast convergence of the other empirical approx. setups:
  - **One-Sample:** $\mathbb{E}\left|\mathsf{W}_1^{(\sigma)}(\hat{P}_n, Q) - \mathsf{W}_1^{(\sigma)}(P, Q)\right| \in O\left(n^{-\frac{1}{2}}\right)$ (GANs)
  - **Two-Sample:** $\mathbb{E}\left|\mathsf{W}_1^{(\sigma)}(\hat{P}_n, \hat{Q}_n) - \mathsf{W}_1^{(\sigma)}(P, Q)\right| \in O\left(n^{-\frac{1}{2}}\right)$

Gaussian-Smoothed OT - Statistical Efficiency

High Level: Fast convergence of empirical approx. under $\mathsf{W}_1^{(\sigma)}$

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

For any $d \geq 1$ and $\sigma > 0$, we have: $\mathbb{E}\mathsf{W}_1^{(\sigma)}(\hat{P}_n, P) \lesssim \left(\frac{c}{\sigma}\right)^{\frac{d}{2}} n^{-\frac{1}{2}}$

Comments:

- True for any subgauassian P (entropic OT assumes compact $\mathrm{supp}(P)$)
- Generalizes to subgauassian noise distributions with monotone density
- Implies fast convergence of the other empirical approx. setups:
  - **One-Sample:** $\mathbb{E}\left|\mathsf{W}_1^{(\sigma)}(\hat{P}_n, Q) - \mathsf{W}_1^{(\sigma)}(P, Q)\right| \in O\left(n^{-\frac{1}{2}}\right)$ (GANs)
  - **Two-Sample:** $\mathbb{E}\left|\mathsf{W}_1^{(\sigma)}(\hat{P}_n, \hat{Q}_n) - \mathsf{W}_1^{(\sigma)}(P, Q)\right| \in O\left(n^{-\frac{1}{2}}\right)$

⊛ **GOT alleviated curse of dimensionality in GAN framework**

Gaussian-Smoothed OT - Summary

$\boxed{\textbf{GOT preserves fav. properties \& alleviates main deficiency}}$

Gaussian-Smoothed OT - Summary

$\boxed{\textbf{GOT preserves fav. properties \& alleviates main deficiency}}$

Next-Generation GANs:

Gaussian-Smoothed OT - Summary

GOT preserves fav. properties & alleviates main deficiency

Next-Generation GANs:

- Attain competitive performance

Gaussian-Smoothed OT - Summary

GOT preserves fav. properties & alleviates main deficiency

Next-Generation GANs:

- Attain competitive performance
- Consume fewer resources

Gaussian-Smoothed OT - Summary

GOT preserves fav. properties & alleviates main deficiency

Next-Generation GANs:

- Attain competitive performance
- Consume fewer resources
- Backed up by (useful) sample complexity guarantees

Gaussian-Smoothed OT - Summary

GOT preserves fav. properties & alleviates main deficiency

Next-Generation GANs:

- Attain competitive performance
- Consume fewer resources
- Backed up by (useful) sample complexity guarantees

Ongoing Work:

Gaussian-Smoothed OT - Summary

$\boxed{\textbf{GOT preserves fav. properties \& alleviates main deficiency}}$

Next-Generation GANs:

- Attain competitive performance
- Consume fewer resources
- Backed up by (useful) sample complexity guarantees

Ongoing Work:

- **Empirical:** GOT-GAN design & test

Gaussian-Smoothed OT - Summary

GOT preserves fav. properties & alleviates main deficiency

Next-Generation GANs:

- Attain competitive performance
- Consume fewer resources
- Backed up by (useful) sample complexity guarantees

Ongoing Work:

- **Empirical:** GOT-GAN design & test
- **Algorithmic:** Enhanced alg. tailored for GOT structure

Gaussian-Smoothed OT - Summary

GOT preserves fav. properties & alleviates main deficiency

Next-Generation GANs:

- Attain competitive performance
- Consume fewer resources
- Backed up by (useful) sample complexity guarantees

Ongoing Work:

- **Empirical:** GOT-GAN design & test
- **Algorithmic:** Enhanced alg. tailored for GOT structure
- **Theoretical:** $\sqrt{n}\mathsf{W}_1^{(\sigma)}(\hat{P}_n, P)$ limiting dist., hypothesis testing, etc.

Recap

- **Generative Adversarial Networks:** SOTA generative models

Recap

- **Generative Adversarial Networks:** SOTA generative models
  - Two perspectives: 'minimax game' and 'min statistical distance'

Recap

- **Generative Adversarial Networks:** SOTA generative models
  - Two perspectives: 'minimax game' and 'min statistical distance'
  - Under 1-Wasserstein metric both coincide

Recap

- **Generative Adversarial Networks:** SOTA generative models
  - Two perspectives: 'minimax game' and 'min statistical distance'
  - Under 1-Wasserstein metric both coincide
  - Wasserstein GANs produce outstanding empirical results

Recap

- **Generative Adversarial Networks:** SOTA generative models
  - Two perspectives: 'minimax game' and 'min statistical distance'
  - Under 1-Wasserstein metric both coincide
  - Wasserstein GANs produce outstanding empirical results
- **Curse of Dimensionality:** Approximate distributions from samples

Recap

- **Generative Adversarial Networks:** SOTA generative models
  - Two perspectives: ‘minimax game’ and ‘min statistical distance’
  - Under 1-Wasserstein metric both coincide
  - Wasserstein GANs produce outstanding empirical results
- **Curse of Dimensionality:** Approximate distributions from samples
  - Empirical approximation under Wasserstein metric is slow $n^{-\frac{1}{d}}$

Recap

- **Generative Adversarial Networks:** SOTA generative models
  - Two perspectives: 'minimax game' and 'min statistical distance'
  - Under 1-Wasserstein metric both coincide
  - Wasserstein GANs produce outstanding empirical results
- **Curse of Dimensionality:** Approximate distributions from samples
  - Empirical approximation under Wasserstein metric is slow $n^{-\frac{1}{d}}$
  - Entropic OT: partial solution but not a metric nor GAN compatible

Recap

- **Generative Adversarial Networks:** SOTA generative models
  - Two perspectives: 'minimax game' and 'min statistical distance'
  - Under 1-Wasserstein metric both coincide
  - Wasserstein GANs produce outstanding empirical results
- **Curse of Dimensionality:** Approximate distributions from samples
  - Empirical approximation under Wasserstein metric is slow $n^{-\frac{1}{d}}$
  - Entropic OT: partial solution but not a metric nor GAN compatible
- **Gaussian-Smoothed OT:** Convolve distributions w/ Gaussians

Recap

- **Generative Adversarial Networks:** SOTA generative models
  - Two perspectives: 'minimax game' and 'min statistical distance'
  - Under 1-Wasserstein metric both coincide
  - Wasserstein GANs produce outstanding empirical results
- **Curse of Dimensionality:** Approximate distributions from samples
  - Empirical approximation under Wasserstein metric is slow $n^{-\frac{1}{d}}$
  - Entropic OT: partial solution but not a metric nor GAN compatible
- **Gaussian-Smoothed OT:** Convolve distributions w/ Gaussians
  - Inherits metric structure from Wasserstein & GAN compatible

Recap

- **Generative Adversarial Networks:** SOTA generative models
  - Two perspectives: ‘minimax game’ and ‘min statistical distance’
  - Under 1-Wasserstein metric both coincide
  - Wasserstein GANs produce outstanding empirical results
- **Curse of Dimensionality:** Approximate distributions from samples
  - Empirical approximation under Wasserstein metric is slow $n^{-\frac{1}{d}}$
  - Entropic OT: partial solution but not a metric nor GAN compatible
- **Gaussian-Smoothed OT:** Convolve distributions w/ Gaussians
  - Inherits metric structure from Wasserstein & GAN compatible
  - Well-behaved function of noise parameter & recovers Wasserstein in limit

Recap

- **Generative Adversarial Networks:** SOTA generative models
  - Two perspectives: 'minimax game' and 'min statistical distance'
  - Under 1-Wasserstein metric both coincide
  - Wasserstein GANs produce outstanding empirical results
- **Curse of Dimensionality:** Approximate distributions from samples
  - Empirical approximation under Wasserstein metric is slow $n^{-\frac{1}{d}}$
  - Entropic OT: partial solution but not a metric nor GAN compatible
- **Gaussian-Smoothed OT:** Convolve distributions w/ Gaussians
  - Inherits metric structure from Wasserstein & GAN compatible
  - Well-behaved function of noise parameter & recovers Wasserstein in limit
  - Fast $n^{-\frac{1}{2}}$ convergence of empirical approximation in all dimensions

Recap

- **Generative Adversarial Networks:** SOTA generative models
  - Two perspectives: 'minimax game' and 'min statistical distance'
  - Under 1-Wasserstein metric both coincide
  - Wasserstein GANs produce outstanding empirical results
- **Curse of Dimensionality:** Approximate distributions from samples
  - Empirical approximation under Wasserstein metric is slow $n^{-\frac{1}{d}}$
  - Entropic OT: partial solution but not a metric nor GAN compatible
- **Gaussian-Smoothed OT:** Convolve distributions w/ Gaussians
  - Inherits metric structure from Wasserstein & GAN compatible
  - Well-behaved function of noise parameter & recovers Wasserstein in limit
  - Fast $n^{-\frac{1}{2}}$ convergence of empirical approximation in all dimensions

Thank you!