Quantifying the Empirical Wasserstein Distance to a Set of Measures: Beating the Curse of Dimensionality

Nian Si

Joint work with Jose Blanchet, Soumyadip Ghosh, and Mark Squillante

NeurIPS 2020

Stanford University

October 22, 2020

Road map

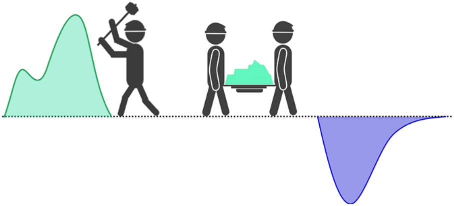
1. Wasserstein Distances and the Curse of Dimensionality
2. Robust Wasserstein Profile Function
3. Duality Results
   - Connections with the the Integral Probability Metric (IPM)
   - Examples
4. Statistical convergence

Wasserstein Distances and the Curse of Dimensionality

Definition of the Wasserstein Distance (earth mover's distance, optimal cost): for any measure P, Q,

$$\mathcal{D}_c(P,Q) = \min_{\pi \in \mathcal{P}(\Omega\times\Omega)} \left\{ \left(\int c(x,w)\pi\left(\mathrm{d}x,\mathrm{d}w\right) \right) \right.$$

$$\left. : \int_{w\in\mathbb{R}^d} \pi\left(\mathrm{d}x,\mathrm{d}w\right) = P\left(\mathrm{d}x\right), \int_{x\in\mathbb{R}^d} \pi\left(\mathrm{d}x,\mathrm{d}w\right) = Q\left(\mathrm{d}w\right) \right\}.$$

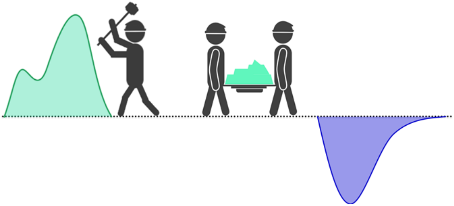


Wasserstein Distances and the Curse of Dimensionality

Definition of the Wasserstein Distance (earth mover's distance, optimal cost): for any measure P, Q,

$$\mathcal{D}_c(P, Q) = \min_{\pi \in \mathcal{P}(\Omega \times \Omega)} \left\{ \left(\int c(x, w) \pi\left(\mathrm{d}x, \mathrm{d}w\right) \right) \right.$$

$$\left. : \int_{w \in \mathbb{R}^d} \pi\left(\mathrm{d}x, \mathrm{d}w\right) = P\left(\mathrm{d}x\right), \int_{x \in \mathbb{R}^d} \pi\left(\mathrm{d}x, \mathrm{d}w\right) = Q\left(\mathrm{d}w\right) \right\}.$$

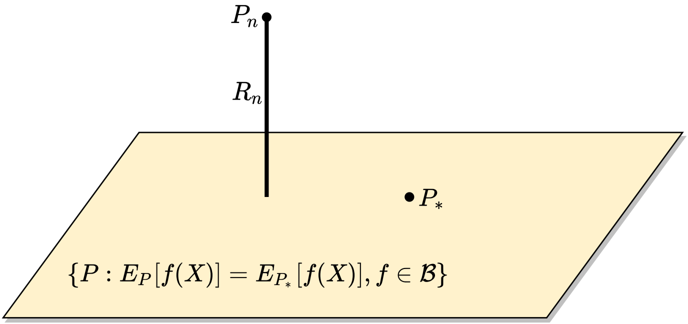


Curse of the dimensionality: $\mathcal{D}_c(P_*, P_n) = O_p\left(n^{-1/(d \vee 2)}\right)$.

- How to explain the good empirical performance, e.g.., Wasserstein GAN?

Robust Wasserstein Profile Function

$$R_n = \inf_{P \in \mathcal{P}(\Omega)} \{\mathcal{D}_c (P, P_n) : \mathbb{E}_P [f (X)] = \mathbb{E}_{P_*} [f (X)] , \text{ for all } f \in \mathcal{B} (\Omega)\}.$$



Duality Results

Theorem (Strong Duality)

Suppose the underlying space Ω is compact and the cost function $c(\cdot,\cdot)$ is a non-negative continuous function with $c(x,x)=0$, for $x \in \Omega$. Then, we have the strong duality

$$\begin{aligned} R_n &:= \inf_{P \in \mathcal{P}(\Omega)} \{\mathcal{D}_c\left(P, P_n\right) : \mathbb{E}_P\left[f\left(X\right)\right] = \mathbb{E}_{P_n}\left[f\left(X\right)\right], \text{ for all } f \in \mathcal{B}\left(\Omega\right)\}. \\ &= \sup_{f \in \mathcal{LB}(\Omega)} \left\{\mathbb{E}_{P_*}\left[f(X)\right] - \mathbb{E}_{P_n}\left[f^c(X)\right]\right\}, \end{aligned}$$

where $f^c(x) = \sup_{z \in \Omega} \{f(z) - c(z,x)\}$ and $\mathcal{LB}(\Omega)$ denotes the linear span generated by $\mathcal{B}(\Omega)$, namely

$$\mathcal{LB}(\Omega) = \left\{ f(\cdot) = \sum_{i=1}^{m} \lambda_i f_i(\cdot) : \{f_i(\cdot)\}_{i=1}^m \subset \mathcal{B}(\Omega), \lambda \in \mathbb{R}^m, \text{ and } m \in \mathbb{Z}_+ \right\}.$$

Duality Results

Theorem (Strong Duality)

Suppose the underlying space Ω *is compact and the cost function* $c(\cdot,\cdot)$ *is a non-negative continuous function with* $c(x,x)=0$, *for* $x\in\Omega$. *Then, we have the strong duality*

$$R_n := \inf_{P\in\mathcal{P}(\Omega)} \{\mathcal{D}_c(P,P_n) : \mathbb{E}_P[f(X)] = \mathbb{E}_{P_n}[f(X)],\ \text{for all } f\in\mathcal{B}(\Omega)\}.$$
$$= \sup_{f\in\mathcal{LB}(\Omega)} \{\mathbb{E}_{P_*}[f(X)] - \mathbb{E}_{P_n}[f^c(X)]\},$$

where $f^c(x) = \sup_{z\in\Omega}\{f(z)-c(z,x)\}$ *and* $\mathcal{LB}(\Omega)$ *denotes the linear span generated by* $\mathcal{B}(\Omega)$, *namely*

$$\mathcal{LB}(\Omega) = \left\{ f(\cdot) = \sum_{i=1}^{m} \lambda_i f_i(\cdot) : \{f_i(\cdot)\}_{i=1}^m \subset \mathcal{B}(\Omega), \lambda\in\mathbb{R}^m,\ \text{and } m\in\mathbb{Z}_+ \right\}.$$

Connections with the the Integral Probability Metric (IPM)

$$\mathrm{IPM}_{\mathcal{F}}(P, P_n) = \sup_{f \in \mathcal{F}} \left| \int f \mathrm{d}P - \int f \mathrm{d}P_n \right|.$$

Connections with the the Integral Probability Metric (IPM)

$$\mathrm{IPM}_{\mathcal{F}}(P, P_n) = \sup_{f \in \mathcal{F}} \left| \int f \mathrm{d}P - \int f \mathrm{d}P_n \right|.$$

- R_n is not a metric in general.

Connections with the the Integral Probability Metric (IPM)

$$\mathrm{IPM}_{\mathcal{F}}(P, P_n) = \sup_{f \in \mathcal{F}} \left| \int f \mathrm{d}P - \int f \mathrm{d}P_n \right|.$$

- R_n is not a metric in general.
- We add a new modeling feature, which is the hypothesis class.

Connections with the the Integral Probability Metric (IPM)

$$\mathrm{IPM}_{\mathcal{F}}(P, P_n) = \sup_{f \in \mathcal{F}} \left| \int f \mathrm{d}P - \int f \mathrm{d}P_n \right|.$$

- R_n is not a metric in general.
- We add a new modeling feature, which is the hypothesis class.
- Our expression for the strong duality uses the combination of both the function f and its c-conjugate f^c in contrast with IPM.

Examples

1. When $\mathcal{B}(\Omega)$ is the space of all 1-Lipschitz functions, $f^c(x) = f(x)$ and R_n reduces to $1-$Wasserstein distance. Then, we recover the Kantorovich-Rubinstein duality result:

$$R_n = \sup_{f \in \mathrm{Lip}_1(\Omega)} \left\{ \mathbb{E}_{P_*}[f(X)] - \mathbb{E}_{P_n}[f(X)] \right\} = \mathcal{D}_1(P_*, P_n).$$

Examples

1. When $\mathcal{B}(\Omega)$ is the space of all 1-Lipschitz functions, $f^c(x) = f(x)$ and R_n reduces to $1-$Wasserstein distance. Then, we recover the Kantorovich-Rubinstein duality result:

$$R_n = \sup_{f \in \mathrm{Lip}_1(\Omega)} \left\{ \mathbb{E}_{P_*}[f(X)] - \mathbb{E}_{P_n}[f(X)] \right\} = \mathcal{D}_1(P_*, P_n).$$

2. Suppose that $\mathcal{B}(\Omega)$ is finite dimensional, such as $\mathcal{B}(\Omega) = \{f_i(x)\}_{i=1}^K$. Then, we have

$$R_n = \sup_{\lambda \in \mathbb{R}^K} \left\{ \mathbb{E}_{P_*}\left[\sum_{i=1}^K \lambda_i f_i(X)\right] - \mathbb{E}_{P_n}\left[\sup_{z \in \Omega} \left\{ \sum_{i=1}^K \lambda_i f_i(z) - c(z, X) \right\}\right] \right\},$$

which recovers the duality result obtained in Blanchet et al. (2019).

Examples

3. Infinite dimensional case: fix linearly independent unit vectors $\theta_1, \ldots, \theta_K \in \mathbb{R}^d$, and consider function class $\mathcal{B}(\Omega) = \cup_{i=1}^K \left\{ f(\theta_i^\top \cdot)|_\Omega : f \in \mathcal{F}_\mathcal{B} \right\}$, where $\mathcal{F}_\mathcal{B}$ collects some 1-dimensional continuous functions, in which case

$$\mathcal{L}\mathcal{B}(\Omega) = \left\{ f(\cdot) = \sum_{i=1}^K \lambda_i f_i(\theta_i^\top \cdot)|_\Omega : \{f_i(\cdot)\}_{i=1}^K \subset \mathcal{F}_\mathcal{B}, \lambda \in \mathbb{R}^K \right\}.$$

Examples

3. Infinite dimensional case: fix linearly independent unit vectors $\theta_1, \ldots, \theta_K \in \mathbb{R}^d$, and consider function class $\mathcal{B}(\Omega) = \cup_{i=1}^K \left\{ f(\theta_i^\top \cdot)|_\Omega : f \in \mathcal{F}_\mathcal{B} \right\}$, where $\mathcal{F}_\mathcal{B}$ collects some 1-dimensional continuous functions, in which case

$$\mathcal{L}\mathcal{B}(\Omega) = \left\{ f(\cdot) = \sum_{i=1}^K \lambda_i f_i(\theta_i^\top \cdot)|_\Omega : \{f_i(\cdot)\}_{i=1}^K \subset \mathcal{F}_\mathcal{B}, \lambda \in \mathbb{R}^K \right\}.$$

Theorem

Following the setting in Example 3 and for $\Omega = \mathbb{R}^d$, we have the strong duality:

$$R_n = \sup_{f \in \mathcal{L}\mathcal{B}(\mathbb{R}^d)} \left\{ \mathbb{E}_{P_*}[f(X)] - \mathbb{E}_{P_n}[f^c(X)] \right\}.$$

Statistical convergence

Theorem

Consider function class $\mathcal{B}(\Omega) = \cup_{i=1}^{K} \left\{ f(\theta_i^\top \cdot)|_\Omega : f \in \mathcal{F}_\mathcal{B} \right\}$ *in Example 3. We assume the space* Ω *is compact and some technical conditions on the function class* $\mathcal{F}_\mathcal{B}$*, we have*

$$nR_n \Rightarrow \sup_{f \in \mathcal{LB}(\Omega)} \left\{ -2H^f - \mathbb{E}_{P_*} \left[\|\nabla_X f(X)\|_2^2 \right] \right\},$$

where $\nabla_x f(x)$ *is the gradient of* $f(\cdot)$ *evaluated at* x *and* H^f *is a Gaussian process indexed by* f *with*

$$H^f \sim \mathcal{N}(0, \mathrm{var}\,(f(X))) \text{ and } \mathrm{cov}(H^{f_1}, H^{f_2}) = \mathrm{cov}\,(f_1(X), f_2(X)).$$

Reference

Si, Nian, Jose Blanchet, Soumyadip Ghosh, and Mark Squillante. "Quantifying the Empirical Wasserstein Distance to a Set of Measures: Beating the Curse of Dimensionality" NeurIPS 2020

Thanks!