A New Partial p-Wasserstein-Based Metric For Comparing Distributions

Sharath Raghvendra, Pouyan Shirzadian, Kaiyi Zhang

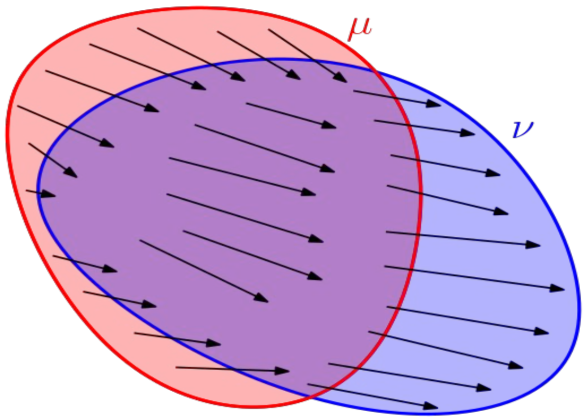
ICML 2024

VIRGINIA TECH.
NC STATE UNIVERSITY

p-Wasserstein Distance

For two probability distributions μ and ν, let γ be a minimum-cost transport plan that transports all the mass from μ to ν. The ***p*-Wasserstein distance** between μ and ν is

$$W_p(\mu, \nu) = \left(\int_{X \times X} \| x - y \|^p \, d\gamma(x, y) \right)^{1/p}.$$



For a parameter $0 \leq \alpha \leq 1$, the **α-partial *p*-Wasserstein distance** between μ and ν is the cost of the minimum-cost transport plan that transports α mass from μ to ν.

Sensitivity of *p*-Wasserstein Distance

✓ **High sensitivity to small geometric differences:** For two distributions μ and $\nu = (1 - \delta)\mu + \delta\nu'$, the p-Wasserstein distance between μ and ν can be as high as $\delta^{1/p}W_p(\mu, \nu')$.

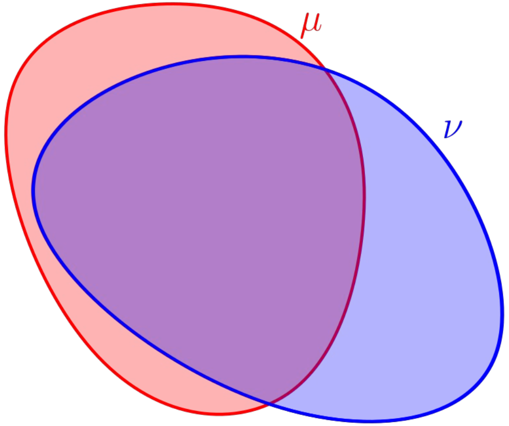
✗ **High sensitivity to outlier:** Adding a noise of δ can cause the distance to be distorted by $\delta^{1/p}$. So, 1% outlier mass ($\delta = 0.01$) can cause the distance to increase by 0.1 for $p = 2$ and by 0.21 for $p = 3$.

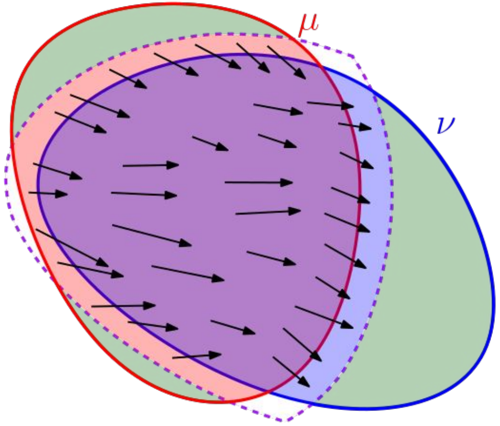
✗ **High sensitivity to sampling discrepancy:** The rate of the convergence of empirical to true p-Wasserstein distance is $n^{-1/2p}$.

Our Distance Function

Given two distributions μ and ν defined over a support X with a unit diameter, the **Robust Partial *p*-Wasserstein distance (*p*-RPW)** between them is the smallest ε for which the cost of $(1 - \varepsilon)$-partial p-Wasserstein distance between μ and ν is at most ε.

$$\Pi_p(\mu, \nu) = \inf\{\varepsilon \geq 0 \mid W_{p,1-\varepsilon}(\mu, \nu) \leq \varepsilon\}$$





Properties of *p*-RPW

Property 1 (Metric Property): The *p*-RPW distance satisfies identity, positivity, symmetry, and the triangle inequality.

Property 2 (Robustness to Outliers): A δ outlier mass can only distort our distance by at most δ.

Property 3 (Faster Convergence): The rate of convergence of the empirical *p*-RPW distance to the true *p*-RPW distance is $n^{-p/(4p-2)}$.

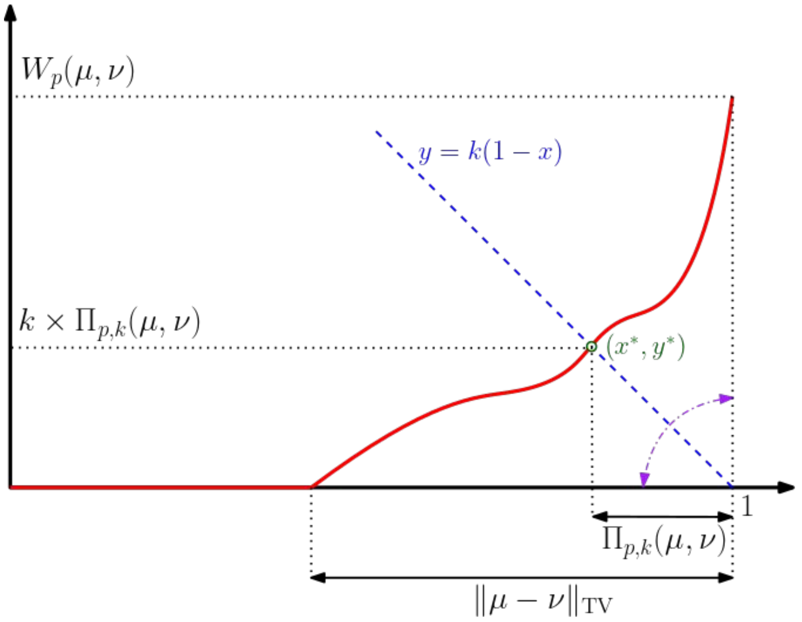
Property 4 (Retaining the Sensitivity of the *p*-Wasserstein Distance): For two distributions μ and ν that differ in δ mass ($||\mu - \nu||_{TV} = \delta$), assuming that the $(1 - \delta/10)$-partial *p*-Wasserstein distance between μ and ν is at least $W_p(\mu, \nu) / 2$, then

$$\Pi_p (\mu, \nu) = \Theta(\min \{\delta, W_p(\mu, \nu)\}).$$

Family of RPW Distances

For a parameter $k \geq 0$, the **(*p*, *k*)-RPW** between two distributions μ and ν is the smallest ε for which the cost of $(1 - \varepsilon)$-partial p-Wasserstein distance between μ and ν is at most $k\varepsilon$.

$$\Pi_{p,k}(\mu, \nu) = \inf\{\varepsilon \geq 0 \mid W_{p,1-\varepsilon}(\mu, \nu) \leq k\varepsilon\}$$



Relation to Well-Known Distances

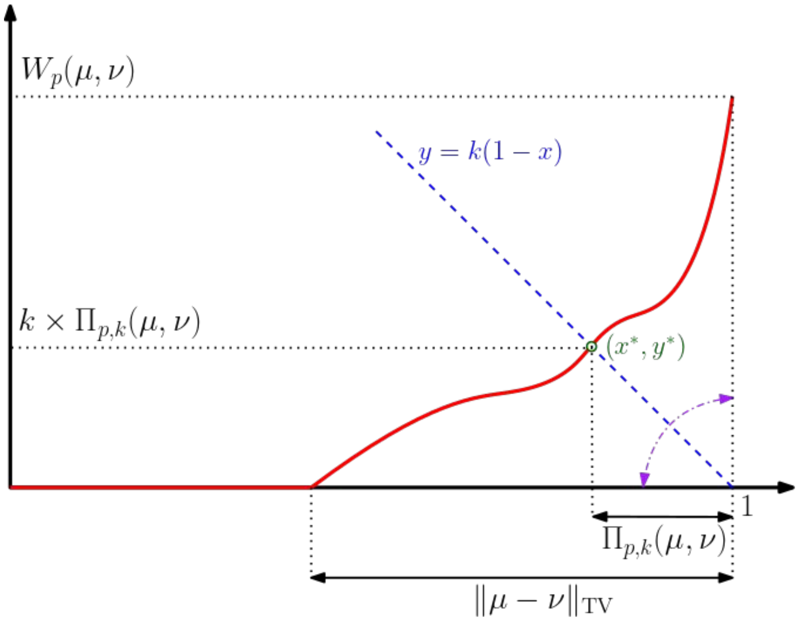
TV distance: When $k = 0$ and for any $p \geq 1$,

$$\Pi_{p,k}(\mu, \nu) = \|\mu - \nu\|_{TV}.$$

***p*-Wasserstein distance:** As $k \to \infty$,

$$\Pi_{p,k}(\mu, \nu) \to W_p(\mu, \nu)/\text{k}.$$

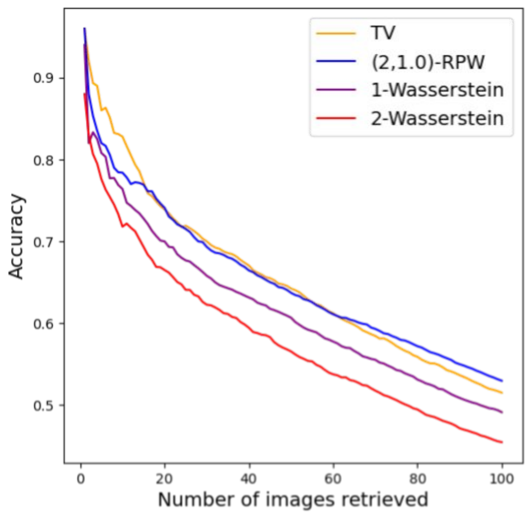
Lévy-Prokhorov distance: For $p = \infty$ and $k = 1$, $\Pi_{p,k}(\mu, \nu)$ is equal to the Lévy-Prokhorov distance between μ and ν.



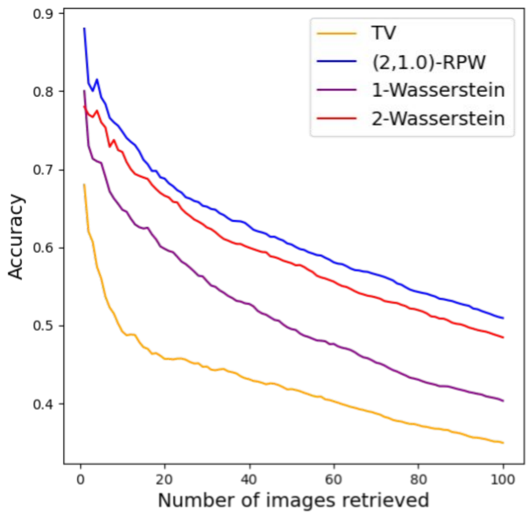
Experiments

Image Retrieval using a labeled dataset of 2k images from MNIST and a query dataset of 50 images.

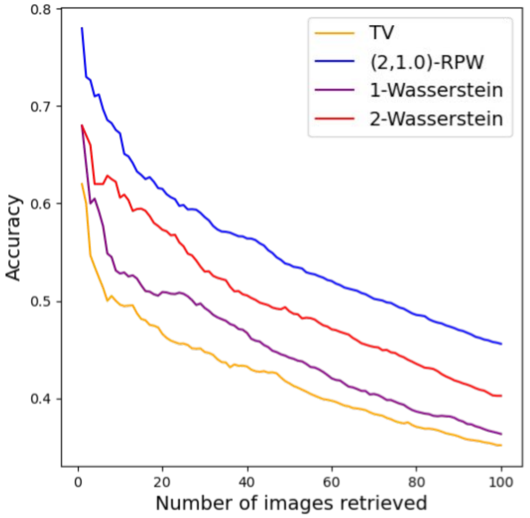
Noise in datasets



Shift in datasets



Noise and shift in datasets



Thanks!