# Outlier-Robust Optimal Transport: Structure, Duality, and Statistical Analysis

Sloan Nietert*, Rachel Cummings†, and Ziv Goldfeld*

*Cornell University, †Columbia University

AISTATS 2022

## Background

The classic Wasserstein distance $\mathrm{W}_p$, defined by

$$\mathrm{W}_p(\mu,\nu) := \Bigg(\inf_{\underset{\uparrow\ \text{set of couplings between } \mu \text{ and } \nu}{\pi\in\Pi(\mu,\nu)}} \int |x-y|^p \, d\pi(x,y)\Bigg)^{\frac{1}{p}},$$

is a popular discrepancy measure between probability measures with many applications in statistics and ML.

## Motivation

Despite its proven utility, $\mathrm{W}_p$ suffers from a sensitivity to outliers, with its strict marginal constraints allowing a small amount of distant mass to contribute greatly to the measured distance[1]. E.g., for any $\varepsilon > 0$,

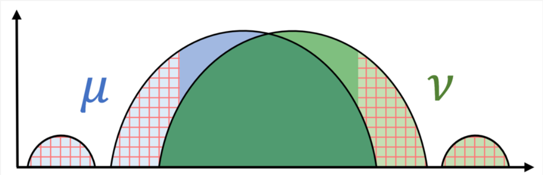
$$\lim_{|x|\to\infty} \mathrm{W}_p(\mu,(1-\varepsilon)\mu + \varepsilon\,\delta_x) = \infty.$$

## Object of Study

**Outlier-robust Wasserstein distance:**

$$\underbrace{\mathrm{W}_p^{\varepsilon}}_{\text{robustness radius}}(\mu,\nu) := \inf_{\substack{\mu',\nu'\in\mathcal{M}_+(\mathbb{R}^d)\\ \mu'\le\mu,\,\nu'\le\nu \\ \mu'(\mathbb{R}^d)=\nu'(\mathbb{R}^d)=1-\varepsilon}} \mathrm{W}_p\left(\frac{\mu'}{1-\varepsilon},\frac{\nu'}{1-\varepsilon}\right), \quad (1)$$

i.e. we remove an $\varepsilon$-fraction of mass from both $\mu$ and $\nu$ (and renormalize) to minimize their OT cost




The gridded light blue and green regions have $\mu$ and $\nu$ mass $\varepsilon$, respectively, and are removed to obtain optimal $\mu'$ and $\nu'$ for $\mathrm{W}_1^{\varepsilon}$

## Population-Limit Robustness Guarantees

**Contamination model:**

$$\|\underbrace{\tilde{\mu}}_{\text{contaminated distributions}} - \overbrace{\mu}^{\text{clean distributions}}\|_{\mathrm{TV}},\ \|\tilde{\nu} - \nu\|_{\mathrm{TV}} \le \varepsilon,$$

**Distributional assumptions:**

$\mu,\nu \in \mathcal{D}$ for $\mathcal{D} \in \{\mathcal{D}_q, \mathcal{D}_2^{\mathrm{cov}}\}$ where

$$\mathcal{D}_q := \left\{\kappa \in \mathcal{P}(\mathbb{R}^d) : \mathbb{E}_\kappa[\|X-x\|^q] \le M \text{ for some } x \in \mathbb{R}^d\right\},$$

$$\mathcal{D}_2^{\mathrm{cov}} := \left\{\kappa \in \mathcal{P}(\mathbb{R}^d) : \Sigma_\kappa \preccurlyeq M^2 I_d\right\}$$

**Minimax risk:**

$\widehat{W}: \mathcal{P}(\mathbb{R}^d)\times\mathcal{P}(\mathbb{R}^d) \to \mathbb{R}$,

$$R_\infty(\mathcal{D},\varepsilon) := \inf_{\widehat{W}} \sup_{\mu,\nu\in\mathcal{D}} \sup_{\substack{\|\tilde{\mu}-\mu\|_{\mathrm{TV}}\le\varepsilon\\ \|\tilde{\nu}-\nu\|_{\mathrm{TV}}\le\varepsilon}} \left|\widehat{W}(\tilde{\mu},\tilde{\nu}) - W_p(\mu,\nu)\right|$$

**Optimality of $W_p^{\varepsilon}$:**

$p < q, \quad \boldsymbol{R_\infty(\mathcal{D}_q,\varepsilon) \asymp M\varepsilon^{1/p-1/q}}$

$p < 2, \quad \boldsymbol{R_\infty(\mathcal{D}_2^{\mathrm{cov}},\varepsilon) \asymp M\sqrt{d}\,\varepsilon^{1/p-1/2}}$

**achieved by** $\boldsymbol{\widehat{W} = W_p^{\varepsilon}}$

## Finite-Sample Robustness Guarantees

**Contamination model:**

$\mathcal{M}^{\mathrm{AC}}(\mu,\nu,\varepsilon)$ — $n$ i.i.d. samples from $\mu$ and $\nu$, $\varepsilon$-fraction arbitrarily corrupted to obtain $\tilde{X}_1,\dots,\tilde{X}_n$ and $\tilde{Y}_1,\dots,\tilde{X}_n$ w/ distribution $P_n$

**Minimax risk:**

Estimator $\widehat{W}_n$ determined by corrupted samples

$$R_n(\mathcal{D},\varepsilon) := \inf_{\widehat{W}_n} \sup_{\mu,\nu\in\mathcal{D}} \sup_{P_n\in\mathcal{M}^{\mathrm{AC}}(\mu,\nu,\varepsilon)} \mathbb{E}_{P_n}\left|\widehat{W}_n - W_p(\mu,\nu)\right|$$

**Optimality of $W_p^{\varepsilon}$:**

In general, $\boldsymbol{R_n(\mathcal{D},\varepsilon) \asymp R_\infty(\mathcal{D},\varepsilon) + \tilde{O}(R_n(\mathcal{D},0))}$

$d > d_0(p,q)$

$p < q,\ \boldsymbol{R_n(\mathcal{D}_q,\varepsilon) \asymp M\varepsilon^{1/p-1/q} + \tilde{O}(n^{-1/d})}$

$p < 2,\ \boldsymbol{R_n(\mathcal{D}_2^{\mathrm{cov}},\varepsilon) \asymp M\sqrt{d}\,\varepsilon^{1/p-1/2} + \tilde{O}(n^{-1/d})}$

**achieved by** $\boldsymbol{\widehat{W}_n = W_p^{\varepsilon}}$

## Duality

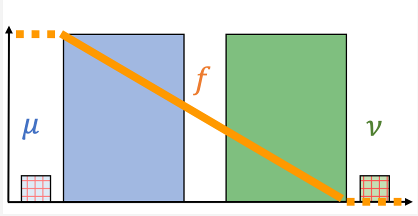
**Kantorovich dual for classic $\mathrm{W}_p$:**

$$\mathrm{W}_p(\mu,\nu)^p = \sup_{\underset{\leftarrow\ \text{continuous, bounded real functions}}{f\in C_b(\mathbb{R}^d)}} \int f\,d\mu + \int \overbrace{f^c}^{\text{c-transform of } f \text{, equal to } -f \text{ when } p=1}\,d\nu$$

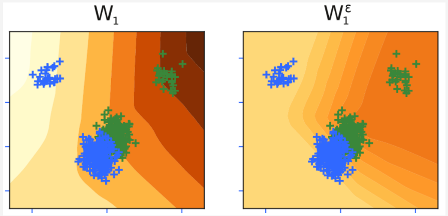
**Dual for our $\mathrm{W}_p^{\varepsilon}$:**

$$(1-\varepsilon)\mathrm{W}_p(\mu,\nu)^p = \sup_{f\in C_b(\mathbb{R}^d)} \int f\,d\mu + \int f^c\,d\nu - 2\varepsilon\|f\|_\infty \quad (2)$$

This elegant dual form is useful for analysis and enables robustification of popular duality-based OT solvers via a simple modification.




1D densities plotted with their optimal potential for the $\mathrm{W}_1^{\varepsilon}$ dual problem




Contour plots for optimal dual potentials to $\mathrm{W}_1$ and $\mathrm{W}_1^{\varepsilon}$ between 2D Gaussian mixtures

## Properties

1. The infimum in (1) and the supremum in (2) are achieved
2. If $f$ is an optimal potential for (2), then there are $\mu' = \mu - \alpha$ and $\nu' = \nu - \beta$ minimizing (1) s.t. $\mathrm{supp}(\alpha) \subseteq \mathrm{argmax}(f)$ and $\mathrm{supp}(\beta) \subseteq \mathrm{argmin}(f)$
   i.e. the max and min level sets of f encode outlier locations

Samples generated by a robustified GAN (left), inspired by the dual formulation (2), alongside samples generated by standard Wasserstein GAN, after training on corrupted MNIST dataset.

1. L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. Unbalanced optimal transport: dynamic and Kantorovich formulations. *Journal of Functional Analysis*, 274(11):3090–3123, 2018.