The classic Wasserstein distance $W_p$, defined by

$$W_p(\mu, \nu) := \left( \inf_{\pi \in \Pi(\mu, \nu)} \int |x - y|^p d\pi(x, y) \right)^{1/p},$$

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

### Motivation

Despite its proven utility, $W_p$ 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. E.g., for any $\epsilon > 0$, $\lim_{|x| \to \infty} W_p(\mu, (1 - \epsilon)\mu + \epsilon \delta_x) = \infty$.

### Object of Study

#### Outlier-robust Wasserstein distance: robustness radius

$$W^\varepsilon_p(\mu, \nu) := \inf_{\mu', \nu' \in \mathcal{M}^\varepsilon(\mathbb{R}^d)} W_p\left(\frac{\mu'}{1 - \varepsilon}, \frac{\nu'}{1 - \varepsilon}\right),$$

$$\mu' \subseteq \mu, \nu' \subseteq \nu, \mu'([d]) = \nu'([d]) = 1 - \varepsilon$$

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

#### Outlier-robust Wasserstein distance: finite sample}

$$W_p^{\varepsilon}(\mu, \nu) := \inf_{\mu', \nu' \in \mathcal{M}^{\varepsilon}(\mathbb{R}^d)} W_p\left(\frac{\mu'}{1 - \varepsilon}, \frac{\nu'}{1 - \varepsilon}\right),$$

$$\mu'([d]) = \nu'([d]) = 1 - \varepsilon$$

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

### Finite-Sample Robustness Guarantees

#### Contamination model: clean distributions

$$\|\tilde{\mu} - \mu\|_{TV}, \|\tilde{\nu} - \nu\|_{TV} \leq \epsilon,$$

contaminated distributions

#### Distributional assumptions:

$$\mu, \nu \in \mathbb{D}$$

where $\mathbb{D} := \{ \mu \in \mathcal{P}(\mathbb{R}^d) : \|\mu - \tilde{\mu}\|_{TV} \leq M \}$

#### Minimax risk:

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

$$R_{\text{inf}}(\mathbb{D}, \varepsilon) := \inf_{\tilde{W}} \sup_{\mu, \nu \in \mathbb{D}} \sup_{\|\tilde{\mu} - \mu\|_{TV} \leq \varepsilon} \|\tilde{W}(\mu, \nu) - W_p(\mu, \nu)\|$$

#### Optimality of $W^\varepsilon$:

$p < q, R_{\text{inf}}(\mathbb{D}, \varepsilon) = M e^{1/p - 1/q}$

achieved by $\tilde{W} = W^\varepsilon_p$