Bounding adapted Wasserstein metrics

Jose Blanchet [1], Martin Larsson [2], Jonghwa Park[2] and Johannes Wiesel [2]

[1] Department of Management Science and Engineering, Standford Univeristy, [2] Department of Mathematical Sciences, Carnegie Mellon University

Introduction to adapted Wasserstein distances

Consider two-period random processes,

$$X = (X_1, X_2) \sim \mu, \quad Y = (Y_1, Y_2) \sim \nu.$$

The **Wasserstein distance** between μ and ν,

$$\mathcal{W}(\mu,\nu) = \inf\{\mathbb{E}^{\pi}[|X-Y|] : \pi \in \mathrm{Cpl}(\mu,\nu)\}$$

is often inadequate when considering time-dependent optimization problems such as optimal stopping problems. Denoting

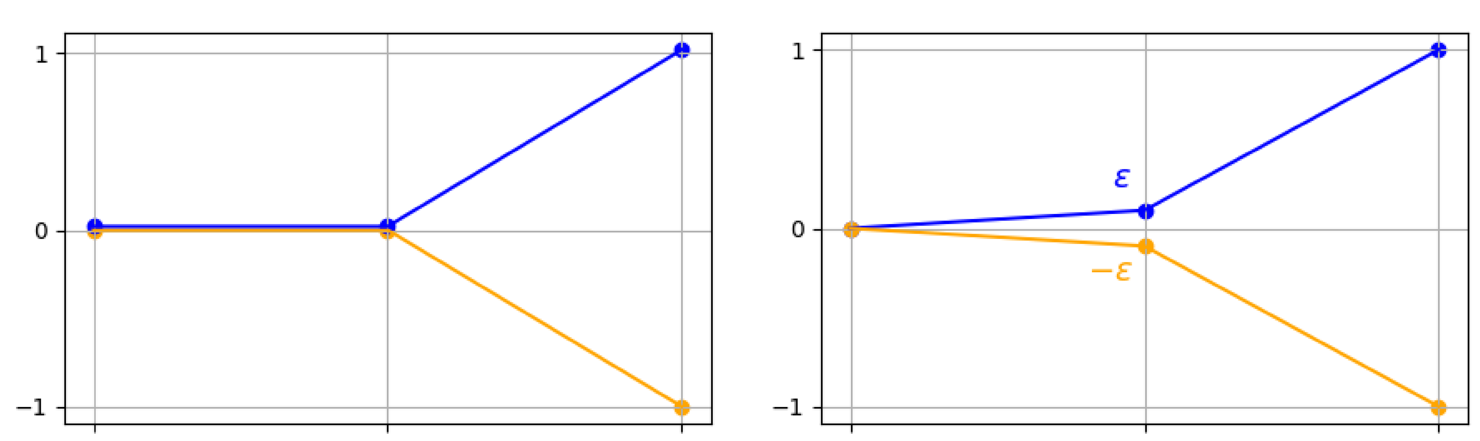
$$\mathrm{Law}(X_1) = \mu_1, \quad \mathrm{Law}(X_2 \mid X_1 = x_1) = \mu_{x_1},$$

it is well-known that the **adapted Wasserstein distance**,

$$\mathcal{AW}(\mu,\nu) = \inf\left\{\mathbb{E}^{\pi_1}\left[|X_1 - Y_1| + \mathcal{W}(\mu_{X_1}, \nu_{Y_1})\right] : \pi_1 \in \mathrm{Cpl}(\mu_1,\nu_1)\right\}$$

address this issue.

The standard example



Paths of $\mu = \frac{1}{2}\delta_{(0,1)} + \frac{1}{2}\delta_{(0,-1)}$ and $\nu = \frac{1}{2}\delta_{(\varepsilon,1)} + \frac{1}{2}\delta_{(-\varepsilon,-1)}$.

They are close in the Wasserstein sense, i.e., $\lim_{\varepsilon\to 0}\mathcal{W}(\mu,\nu) = 0$, but they are not close in the adapted Wasserstein sense, i.e., $\lim_{\varepsilon\to 0}\mathcal{AW}(\mu,\nu) = 1 \neq 0$.

Motivation & Goal

Besides $\mathcal{W} \le \mathcal{AW}$, relatively little is known about the **metric space** $(\mathcal{P}_1(\mathbb{R}^2), \mathcal{AW})$. We aim to answer the following question;

> Can we revert $\mathcal{W} \le \mathcal{AW}$, i.e., can we find a nice function $f : \mathbb{R} \to \mathbb{R}$ such that $\mathcal{AW} \le f(\mathcal{W})$?

The answer is no in general. However, it follows from the classical theory that on a $\mathcal{AW}_1$-relatively compact set $K \subseteq \mathcal{P}_1(\mathbb{R}^2)$, we have that for all $\sigma > 0$,

$$\mathcal{AW}(\mu,\nu) \le C\mathcal{W}(\mu,\nu) + \sigma, \quad \mu,\nu \in K$$

where C depends on K and σ. Our contribution is to characterize the dependence of C on K and σ as follows;

- For any $0 < \sigma < R$,
$$\mathcal{AW}(\mu,\nu) \lesssim \frac{R}{\sigma}\mathcal{W}(\mu,\nu) + g(R,K) + h(\sigma,K).$$
- A function g characterizes the tail behavior outside a Euclidean ball of radius R and $\lim_{R\to\infty} g(R,K) = 0$
- A function h is derived from the modulus of continuity and $\lim_{\sigma\to 0} h(\sigma,K) = 0$.

Smooth adapted Wasserstein distances

The **smooth adapted Wasserstein distance** as

$$\mathcal{AW}^{(\sigma)}(\mu,\nu) = \mathcal{AW}(\mu * \mathcal{N}_\sigma, \nu * \mathcal{N}_\sigma) \text{ where } \mathcal{N}_\sigma = \mathcal{N}(0, \sigma^2 \mathrm{I}).$$

Theorem (Blanchet, Larsson, Park, Wiesel, '24). For any $\sigma, R > 0$,

$$\mathcal{AW}^{(\sigma)}(\mu,\nu) \lesssim \frac{R}{\sigma}\mathcal{W}(\mu,\nu) + \int_{\{|x|\ge R\}} |x| \, (\mu * \mathcal{N}_\sigma + \nu * \mathcal{N}_\sigma)(dx).$$

Corollary (Blanchet, Larsson, Park, Wiesel, '24). Assume $\mu,\nu \in \mathcal{P}_q(\mathbb{R}^2)$ for $q > 1$ and $0 < \tau < \sigma$.

$$\mathcal{AW}^{(\sigma)}(\mu,\nu) \lesssim \left(\sigma + \left(\mathbb{E}^{\mu}[|X|^q]\right)^{1/q} + \left(\mathbb{E}^{\nu}[|Y|^q]\right)^{1/q}\right)\left(\frac{\mathcal{W}^{(\tau)}(\mu,\nu)}{\sqrt{\sigma^2 - \tau^2}}\right)^{1-1/q}.$$

Corollary (Blanchet, Larsson, Park, Wiesel, '24). Assume $\mu \in \mathcal{P}_q(\mathbb{R}^2)$ for $q > 4$. If $\widehat{\mu}_n$ is the empirical measure of n i.i.d samples of μ,

$$\mathbb{E}[\mathcal{AW}^{(\sigma)}(\mu, \widehat{\mu}_n)] \lesssim n^{-\frac{1}{2}+\frac{1}{2q}}.$$

Smoothing and the modulus of continuity

The **modulus of continuity** of μ is given as

$$\omega_\mu(\sigma) = \sup\left\{\mathbb{E}^{\pi_1}\left[\mathcal{W}(\mu_{X_1}, \mu_{X_1'})\right] : \pi_1 \in \mathrm{Cpl}(\mu_1,\mu_1), \mathbb{E}^{\pi_1}[|X_1 - X_1'|] < \sigma\right\}.$$

Proposition (Eder, '19). A set $K \subseteq \mathcal{P}_1(\mathbb{R}^2)$ is $\mathcal{AW}$-relatively compact if and only if K is $\mathcal{W}$-relatively compact and $\lim_{\sigma\searrow 0}\sup_{\mu\in K}\omega_\mu(\sigma) = 0$.

Theorem (Blanchet, Larsson, Park, Wiesel, '24). For any $\sigma > 0$,

$$\mathcal{AW}(\mu, \mu * \mathcal{N}_\sigma) \lesssim \sigma + \omega_\mu(\sigma).$$

Upperbounds of $\mathcal{AW}$ in terms of $\mathcal{W}$

From the triangle inequality,

$$\mathcal{AW}(\mu,\nu) \le \mathcal{AW}^{(\sigma)}(\mu,\nu) + \mathcal{AW}(\mu, \mu * \mathcal{N}_\sigma) + \mathcal{AW}(\nu, \nu * \mathcal{N}_\sigma).$$

Corollary (Blanchet, Larsson, Park, Wiesel, '24). Assume $K \subseteq \mathcal{P}_1(\mathbb{R}^2)$ is $\mathcal{AW}$-relatively compact. For $\mu,\nu \in K$ and $0 < \sigma < R$,

$$\mathcal{AW}(\mu,\nu) \lesssim \frac{R}{\sigma}\mathcal{W}(\mu,\nu) + \sup_{\gamma\in K}\int_{\{|x|\ge R\}} |x| \, \gamma(dx) + \sup_{\gamma\in K}(\sigma + \omega_\gamma(\sigma)).$$

Corollary (Blanchet, Larsson, Park, Wiesel, '24). Assume μ,ν are supported on a bounded set $F \subseteq \mathbb{R}^2$ with α-Hölder continuous kernels, i.e. $\mathcal{W}(\mu_x, \mu_y) \le L|x-y|^\alpha$. Then

$$\mathcal{AW}(\mu,\nu) \lesssim \mathcal{W}(\mu,\nu)^{\alpha/(\alpha+1)}.$$

Comparison of topologies

The decay of smoothing parameters σ_n determines the smooth adapted Wasserstein topology as illustrated in the following diagram.

$$\begin{tikzcd}
\mathcal{W}(\mu,\mu_n) \to 0 \arrow[d, Leftrightarrow] \arrow[r, bend left, "\text{if } \mathcal{W}(\mu,\mu_n)/\sigma_n \to 0"] & \mathcal{AW}^{(\sigma_n)}(\mu,\mu_n) \to 0 \arrow[l, Rightarrow] \arrow[r, bend right, "\text{if } \omega_{\mu_n}(\sigma_n)\to 0"'] & \mathcal{AW}(\mu,\mu_n) \to 0 \arrow[l, Rightarrow] \\
\mathcal{W}^{(\sigma)}(\mu,\mu_n) \to 0 \arrow[d, Leftrightarrow] & & \\
\mathcal{W}^{(\sigma_n)}(\mu,\mu_n) \to 0 \arrow[d, Leftrightarrow] & & \\
\mathcal{AW}^{(\sigma)}(\mu,\mu_n) \to 0 & &
\end{tikzcd}$$

If σ_n decays slow enough, then $\mathcal{AW}^{(\sigma_n)}(\mu,\mu_n)$ mirrors $\mathcal{W}(\mu,\mu_n)$. If σ_n decays fast enough, i.e. $\omega_{\mu_n}(\sigma_n) \to 0$, then $\mathcal{AW}^{(\sigma_n)}(\mu,\mu_n)$ and $\mathcal{AW}(\mu,\mu_n)$ converge towards the same value.

Generalization

- The same results hold for any finite time processes on Eudlidean spaces.
- Possible to choose any distributions as smoothing kernels as long as it satisfies the mild growth condition, e.g., gaussian, compactly supported measures.

References

[1] J. Backhoff-Veraguas, D. Bartl, M. Beiglböck, and M. Eder. All adapted topologies are equal. *Probability Theory and Related Fields*, 178:1125–1172, 2020.

[2] J. Blanchet, M. Larsson, J. Park, and J. Wiesel. Bounding adapted wasserstein metrics. *arXiv preprint arXiv:2407.21492*, 2024.

[3] M. Eder. Compactness in adapted weak topologies. *arXiv preprint arXiv:1905.00856*, 2019.

[4] R. Sadhu, Z. Goldfeld, and K. Kato. Limit distribution theory for the smooth 1-Wasserstein distance with applications. *arXiv preprint arXiv:2107.13494*, 2021.