Bounding adapted Wasserstein metrics

Johannes Wiesel

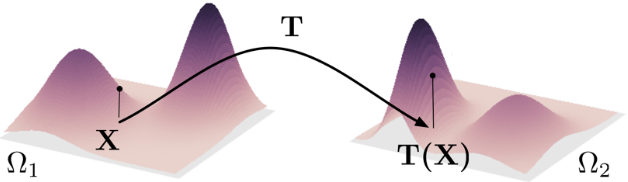
Department of Mathematics
Carnegie Mellon University

based on joint works with
Jose Blanchet, Martin Larsson and Jonghwa Park

Bachelier Finance Society One World Seminar, October 2024

Adapted Wasserstein metric:
a two-period framework

Optimal transport and Wasserstein metric



$$W(\mu,\nu) = \inf\left\{ \int \|x-y\|\,\pi(dx,dy) \colon \pi \in \Pi(\mu,\nu) \right\},$$

where $\Pi(\mu,\nu)$ is the set of couplings between μ and ν.

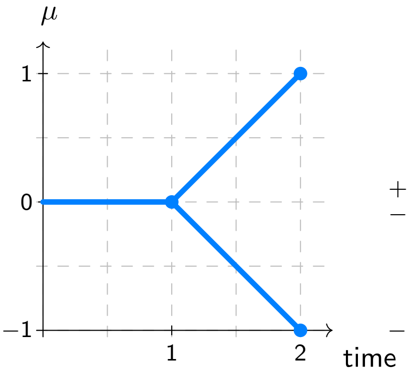
A two-period framework

- Assume now we are observing a two-period stochastic process $X = (X_1, X_2)$.
- Ignoring time, we can see it as an $\mathbb{R}^2$-valued random variable and compute Wasserstein distances in $\mathcal{P}(\mathbb{R}^2)$, i.e.

$$W(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int |x_1 - y_1| + |x_2 - y_2| \, \pi(dx, dy).$$

- So what's the big deal?

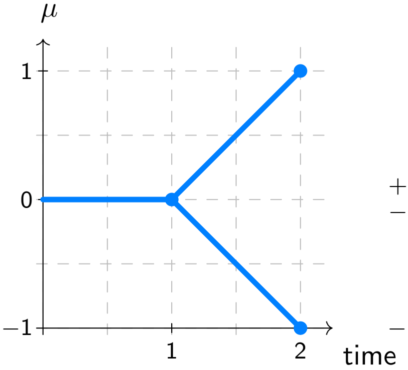
The standard example



i.e.

$$\mu = \frac{1}{2}\delta_{(0,1)} + \frac{1}{2}\delta_{(0,-1)},$$
$$\mu^{\varepsilon} = \frac{1}{2}\delta_{(\varepsilon,1)} + \frac{1}{2}\delta_{(-\varepsilon,-1)}.$$

The standard example



Note that

$$\lim_{\varepsilon \downarrow 0} W(\mu, \mu^\varepsilon) = 0, \text{ i.e. "atoms get closer".}$$

<u>but</u>: $\mu^\varepsilon(\cdot | X_1 = \pm\varepsilon) = \delta_{\pm 1} \not\to \mu(\cdot | X_1 = 0)$.

Adapted Wasserstein metric

The "adapted" Wasserstein distance respects the time-structure of the transport:

$$AW(\mu,\nu) := \inf_{\pi^1 \in \Pi(\mu^1,\nu^1)} \int |x_1 - y_1| + W(\mu_{x_1},\nu_{y_1})\, \pi^1(dx_1, dy_1)$$

- couple first μ^1 and ν^1 $\Rightarrow$ are first marginals close?
- based on this coupling, couple $\mu_{x_1}(\cdot) = \mu(\cdot|X_1 = x_1)$ and $\nu_{y_1}(\cdot) = \nu(\cdot|Y_1 = y_1)$ $\Rightarrow$ are conditional laws close?
- always true: $W(\mu,\nu) \le AW(\mu,\nu)$.
- Standard example:
  $\lim_{\varepsilon\downarrow 0} W(\mu,\mu^\varepsilon) = 0$, $\lim_{\varepsilon\downarrow 0} AW(\mu,\mu^\varepsilon) = 1 \neq 0$.

Adapted weak topology: a little bit of history

Versions of the adapted Wasserstein topology have been (re-)discovered in many different contexts:

- Aldous '81: extended weak convergence
- Hoover, Keisler '84: Model theory
- Rüschendorf '85: Markov-constructions
- Hellwig '96: information topology
- Pflug-Pichler '12: nested distance for multistage optimistation
- Lasalle '18: (bi-)causal couplings
- Acciaio, Backhoff, Beiglböck, ... '18: filtered processes
- Bonnier, Liu, Oberhauser '20: higher rank signatures
- ...

Adapted weak topology

Theorem (Backhoff, Bartl, Beiglböck, Eder, PTRF '20)

All topologies listed above are equal to the topology induced by AW. It is the coarsest topology, that makes the functional

$$\mu \mapsto \sup_{\tau \in \{1,2\}} \int x_\tau \, \mu(dx)$$

continuous.

- many further results known: completion of $(\mathcal{P}(\mathbb{R}^2), AW)$, geodesics, pre-compact sets, barycenters, ...

Adapted Wasserstein metric

For applications e.g. in mathematical finance, statistics, . . . , we are mostly interested in comparisons for the metric AW, not the adapted weak topology:

Theorem (Backhoff, Bartl, Beiglböck, Eder, F&S '20)

Many optimization problems in mathematical finance (utility maximization, indifference pricing, superhedging, . . .)

$$\mu \mapsto inf_\alpha \int f(x,\alpha)\,\mu(dx)$$

are Lipschitz continuous wrt. AW.

Problems with adapted Wasserstein metric

But: dealing with AW is hard!

- AW is difficult to compute/approximate: nested optimization problem.
- Not a lot of (useful) characterisations of optimizers, like Brenier's theorem available.
- W-consistent estimators are usually not AW-consistent: e.g. empirical measure vs adapted empirical measure (Bartl, Backhoff, Beiglböck, Wiesel, AAP, '22).
- relatively little is known about the metric space $(\mathcal{P}(\mathbb{R}^2), AW)$: $W \leq AW$ but not much else...
- Aim of this talk: shed some light on this!

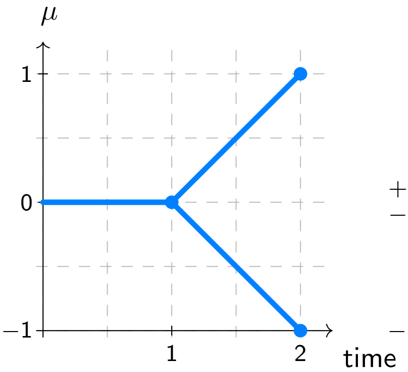
A (naive) goal

- Can we revert $W \leq AW$, i.e. can we find a "nice" function $f : \mathbb{R} \to \mathbb{R}$, such that $AW \leq f(W)$?
- If true, then big advantage: know much more about Wasserstein distance (e.g. in statistics, optimization)
- Without restrictions, bounds of type $AW \leq f(W)$ have to be useless: in example above

$$1 = \lim_{\varepsilon \to 0} AW(\mu^\varepsilon, \mu), \quad \lim_{\varepsilon \to 0} W(\mu^\varepsilon, \mu) = 0.$$

Ingredients

The standard example

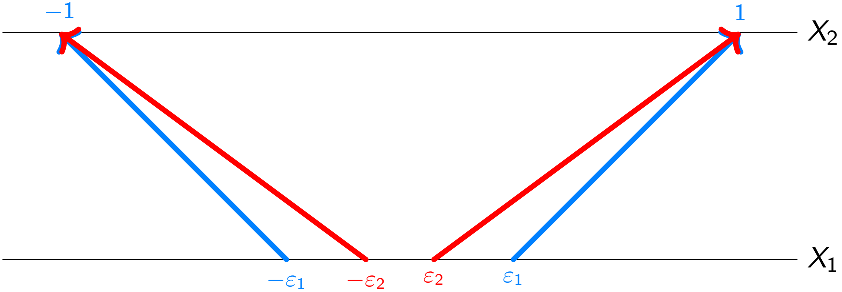


$$1 = \lim_{\varepsilon \to 0} AW(\mu^\varepsilon, \mu), \quad \lim_{\varepsilon \to 0} W(\mu^\varepsilon, \mu) = 0.$$

Question: why are things going wrong here?

Standard example revisited: regularity of conditional laws

Key point: regularity of $\mu^\varepsilon(\cdot|X_1)$ gets worse for $\varepsilon \to 0$



- Analogy: think of $f_\varepsilon : \mathbb{R} \to \mathbb{R}$ continuous with $f_\varepsilon(-\varepsilon) := -1, f_\varepsilon(\varepsilon) := 1$.
- f_ε is not equicontinuous at 0: $f_\varepsilon(\varepsilon) - f_\varepsilon(-\varepsilon) = 2!$

AW-relatively compact sets

Theorem (Eder, '19)

A set $K \subseteq \mathcal{P}(\mathbb{R}^2)$ is AW-relatively compact if and only if K is W-relatively compact and

$$\lim_{\delta \to 0} \sup_{\mu \in K} \omega_\mu(\delta) = 0 \tag{1}$$

where $\omega_\mu : (0, \infty) \to (0, \infty)$ is the modulus of continuity

$$\omega_\mu(\delta) = \sup \{\mathbb{E}[W(\mu(\cdot|X_1), \mu(\cdot|Y_1)] : X, Y \sim \mu, \mathbb{E}[|X_1 - Y_1|] < \delta\}.$$

AW-relatively compact sets

Theorem (Eder, '19)

A set $K \subseteq \mathcal{P}(\mathbb{R}^2)$ is AW-relatively compact if and only if K is W-relatively compact and

$$\lim_{\delta\to 0}\sup_{\mu\in K}\omega_\mu(\delta) = 0 \tag{1}$$

where $\omega_\mu : (0,\infty) \to (0,\infty)$ is the modulus of continuity

$$\omega_\mu(\delta) = \sup\left\{\mathbb{E}[W(\mu(\cdot|X_1), \mu(\cdot|Y_1)] : X, Y \sim \mu, \mathbb{E}[|X_1 - Y_1|] < \delta\right\}.$$

$\Rightarrow K = \{\mu^\varepsilon : \varepsilon > 0\}$ in standard example is not *AW*-precompact!

Examples of AW-relatively compact sets

However, many classes of measures are AW-relatively compact!

Definition

$\mu \in \mathcal{P}(\mathbb{R}^2)$ has α-Hölder kernels, if

$$W(\mu(\cdot|X_1 = x_1), \mu(\cdot|X_1 = y_1)) \leq L|x_1 - y_1|^\alpha.$$

Examples of *AW*-relatively compact sets

However, many classes of measures are *AW*-relatively compact!

Definition

$\mu \in \mathcal{P}(\mathbb{R}^2)$ has α-Hölder kernels, if

$$W(\mu(\cdot|X_1 = x_1), \mu(\cdot|X_1 = y_1)) \leq L|x_1 - y_1|^\alpha.$$

Lemma (Blanchet, Larsson, Park, W. '24)

A measure $\mu \in \mathcal{P}(\mathbb{R}^2)$ has Lipschitz kernels (i.e. $\alpha = 1$) if one of the following holds:

- *it is supported on finitely many points.*
- *it has a Lipschitz-continuous density, which is bounded away from zero.*
- *$X_2 = F(X_1, \varepsilon)$ where F is Lipschitz and ε is independent noise.*
- *$X_1 = X + \varepsilon_1, X_2 = Y + \varepsilon_2$, where $(\varepsilon_1, \varepsilon_2)$ is independent noise.*

The smooth adapted Wasserstein distance

Definition

Define the smooth adapted Wasserstein distance

$$AW^{\sigma}(\mu,\nu) := AW(\mu * \mathcal{N}(0,\sigma^2 I), \nu * \mathcal{N}(0,\sigma^2 I))$$

for $\sigma > 0$.

The smooth adapted Wasserstein distance

Definition

Define the smooth adapted Wasserstein distance

$$AW^\sigma(\mu,\nu) := AW(\mu * \mathcal{N}(0,\sigma^2 I), \nu * \mathcal{N}(0,\sigma^2 I))$$

for $\sigma > 0$.

- the smooth (non-adapted) Wasserstein distance $W^\sigma := W(\mu * \mathcal{N}(0,\sigma^2 I), \nu * \mathcal{N}(0,\sigma^2 I))$ is well-studied in the (statistics) literature; we have $W^\sigma \leq W$.
- $AW^\sigma(\mu,\nu)$ and its statistical properties introduced only very recently in Blanchet, W., 2xZhang '24 and Hou '24.
- AW^σ is a metric.

Main results

Putting things together

- Recall our aim: want to find "nice" $f : \mathbb{R} \to \mathbb{R}$ such that $AW \leq f(W)$.
- Use the basic estimate

$$AW(\mu,\nu) \leq AW(\mu,\mu^\sigma) + AW(\mu^\sigma,\nu^\sigma) + AW(\nu,\nu^\sigma)$$
$$\leq 2AW(\mu,\mu^\sigma) \vee AW(\nu,\nu^\sigma) + AW^\sigma(\mu,\nu).$$

  where $\mu^\sigma := \mu * \mathcal{N}(0,\sigma^2 I)$.
- Now bound terms on the rhs separately.

Main results

Theorem (Blanchet, Larsson, Park, W. '24)

For any $\sigma, R > 0$ we have

$$AW^{\sigma}(\mu,\nu) \leq C\left(R\frac{W(\mu,\nu)}{\sigma} + \int_{\{|x|\geq R\}} |x|^p(\mu^{\sigma}+\nu^{\sigma})(dx)\right).$$

i.e. smoothed AW can be bounded by W + tail bound!

Main results

Corollary (Blanchet, Larsson, Park, W. '24)

Assume $\mu, \nu \in \mathcal{P}_q(\mathbb{R}^2)$ *for* $q > 1$. *For any* $\sigma > \sigma_0 > 0$ *we have*

$$AW^\sigma(\mu, \nu) \leq C \left(\frac{W^{\sigma_0}(\mu, \nu)}{\sqrt{\sigma^2 - \sigma_0^2}} \right)^{1-1/q},$$

where we recall $W^{\sigma_0} \leq W$.

Main results

Corollary (Blanchet, Larsson, Park, W. '24)

Assume $\mu, \nu \in \mathcal{P}_q(\mathbb{R}^2)$ for $q > 1$. For any $\sigma > \sigma_0 > 0$ we have

$$AW^{\sigma}(\mu, \nu) \leq C \left(\frac{W^{\sigma_0}(\mu, \nu)}{\sqrt{\sigma^2 - \sigma_0^2}} \right)^{1-1/q},$$

where we recall $W^{\sigma_0} \leq W$.

Corollary (Blanchet, Larsson, Park, W. '24)

If μ_n is the empirical measure of n i.i.d. samples from $\mu \in \mathcal{P}_q(\mathbb{R}^2)$ for some $q > 4$, then

$$\mathbb{E}[AW^{\sigma}(\mu, \mu_n)] \leq Cn^{-(q-1)/(2q)},$$

where C depends on $\sigma, d, q, M_q(\mu)$.

Main results (2)

Theorem (Blanchet, Larsson, Park, W. '24)

For any $\sigma > 0$ we have

$$AW(\mu^\sigma, \mu) \leq C(\sigma + \omega_\mu(\sigma)).$$

where we recall

$$\omega_\mu(\sigma) = \sup \left\{ \mathbb{E}[W(\mu(\cdot|X_1), \mu(\cdot|Y_1))] : X, Y \sim \mu, \mathbb{E}[|X_1 - Y_1|] < \sigma \right\}.$$

Main results (2)

Theorem (Blanchet, Larsson, Park, W. '24)

For any $\sigma > 0$ we have

$$AW(\mu^\sigma, \mu) \leq C(\sigma + \omega_\mu(\sigma)).$$

where we recall

$$\omega_\mu(\sigma) = \sup \left\{ \mathbb{E}[W(\mu(\cdot|X_1), \mu(\cdot|Y_1)] : X, Y \sim \mu, \mathbb{E}[|X_1 - Y_1|] < \sigma \right\}.$$

and

$$\lim_{\sigma \to 0} \sup_{\mu \in K} \omega_\mu(\sigma) = 0 \qquad (2)$$

for AW-precompact sets K by Eder '19.

Now stir everything well...

Corollary (Blanchet, Larsson, Park, W. '24)

For all $\sigma, R > 0$ *we have*

$$
\begin{aligned}
AW(\mu,\nu) &\leq AW^{\sigma}(\mu,\nu) + 2AW(\mu,\mu^{\sigma}) \vee AW(\nu,\nu^{\sigma}) \\
&\lesssim \left(R\frac{W(\mu,\nu)}{\sigma} + \int_{\{|x|\geq R\}} |x|^p(\mu^{\sigma}+\nu^{\sigma})(dx) \right) \\
&\quad + \sigma + \omega_{\mu}(\sigma) \vee \omega_{\nu}(\sigma).
\end{aligned}
$$

Now stir everything well...

Corollary (Blanchet, Larsson, Park, W. '24)

For all $\sigma, R > 0$ *we have*

$$\begin{aligned} AW(\mu,\nu) &\leq AW^{\sigma}(\mu,\nu) + 2AW(\mu,\mu^{\sigma}) \vee AW(\nu,\nu^{\sigma}) \\ &\lesssim \left(R\frac{W(\mu,\nu)}{\sigma} + \int_{\{|x|\geq R\}} |x|^p(\mu^{\sigma}+\nu^{\sigma})(dx) \right) \\ &\quad + \sigma + \omega_{\mu}(\sigma) \vee \omega_{\nu}(\sigma). \end{aligned}$$

Corollary (Blanchet, Larsson, Park, W. '24)

Let $F \subseteq \mathbb{R}^2$ *be bounded, let* $\mu, \nu \in \mathcal{P}(F)$ *with* α*-Hölder continuous kernels. Then*

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

Comparison of topologies

Comparison of topologies

$$W(\mu, \mu_n) \to 0 \Longleftarrow AW(\mu, \mu_n) \to 0$$

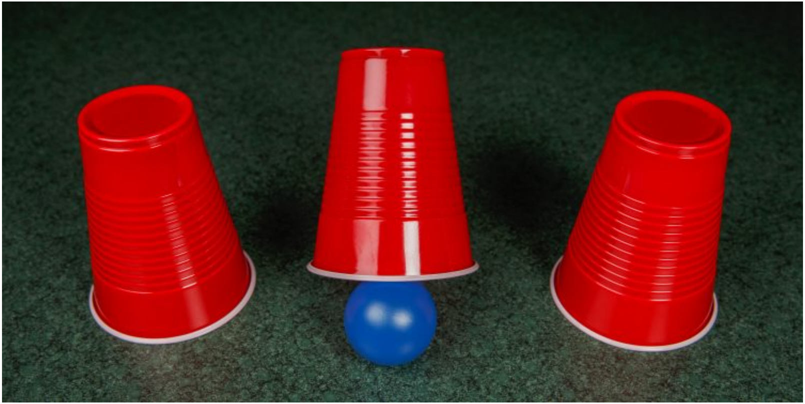
Comparison of topologies

$$\begin{array}{ccc} W(\mu, \mu_n) \to 0 & \Longleftarrow & AW(\mu, \mu_n) \to 0 \\ \Updownarrow & & \\ W^{\sigma}(\mu, \mu_n) \to 0 & & \\ \Updownarrow & & \\ W^{\sigma_n}(\mu, \mu_n) \to 0 & & \end{array}$$

Comparison of topologies

$$
\begin{array}{ccccc}
W(\mu,\mu_n) \to 0 & \Longleftarrow & {\color{red} AW^{\sigma_n}(\mu,\mu_n) \to 0} & \Longleftarrow & AW(\mu,\mu_n) \to 0 \\
\Updownarrow & & & & \\
W^{\sigma}(\mu,\mu_n) \to 0 & & & & \\
\Updownarrow & & & & \\
W^{\sigma_n}(\mu,\mu_n) \to 0 & & & &
\end{array}
$$

Comparison of topologies



Comparison of topologies

if $W(\mu,\mu_n)/\sigma_n \to 0$

$$W(\mu,\mu_n) \to 0 \Longleftarrow AW^{\sigma_n}(\mu,\mu_n) \to 0 \Longleftarrow AW(\mu,\mu_n) \to 0$$

if $\omega_{\mu_n}(\sigma_n) \to 0$

$$\Updownarrow$$

$$W^{\sigma}(\mu,\mu_n) \to 0$$

$$\Updownarrow$$

$$W^{\sigma_n}(\mu,\mu_n) \to 0$$

$$\Updownarrow$$

$$AW^{\sigma}(\mu,\mu_n) \to 0$$

Thank You!

preprint available at

https://sites.google.com/view/johannes-wiesel

The standard example one more time

Lemma

Let $\mu_n = \frac{1}{2}\delta_{(1+\varepsilon_n,2)} + \frac{1}{2}\delta_{(1-\varepsilon_n,0)}$ *and* $\mu = \frac{1}{2}\delta_{(1,2)} + \frac{1}{2}\delta_{(1,0)}$. *As* $\varepsilon_n \to 0$, $W(\mu,\mu_n) \to 0$ *and* $AW(\mu,\mu_n) \to 1 \neq 0$. *We can show that*

- $\varepsilon_n/\sigma_n \to 0$ *iff* $W(\mu_n,\mu)/\sigma_n \to 0$ *iff* $AW^{\sigma_n}(\mu_n,\mu) \to 0$.
- $\varepsilon_n/\sigma_n \to \infty$ *iff* $\omega_{\mu_n}(\sigma_n) = (\sigma_n/\varepsilon_n) \wedge 2 \to 0$ *iff* $AW(\mu_n^{\sigma_n},\mu_n) \to 0$.