On a T_1 Transport inequality for the adapted Wasserstein distance

Jonghwa Park

Department of Mathematical Sciences, Carnegie Mellon University

The weak topology

How can we define a topology on the space of probability measures ?

- The weak topology is the smallest topology for which
$$\mathbb{P} \mapsto \mathbb{E}^{\mathbb{P}}[\varphi(X)]$$
is continuous for all continuous and bounded φ.
- The Wasserstein distance metrizes the weak topology:
$$\mathcal{W}_1(\mathbb{P},\mathbb{Q}) = \inf\{\mathbb{E}^{\pi}[|X-Y|] : \pi \text{ has marginals } \mathbb{P},\mathbb{Q}\}.$$

A topology for stochastic processes

$\mathbb{P}$ is a law of a stochastic process $X = (X_1, X_2, \ldots, X_N)$. Ideally, we seek a topology for which the map
$$\mathbb{P} \mapsto \nu(\mathbb{P})$$
is continuous for time-dependent ν such as

- (optimal stopping problem) $\nu(\mathbb{P}) = \inf\{\mathbb{E}^{\mathbb{P}}[\varphi(X_\tau)] : \text{ stopping time } \tau\}$,
- (superhedging error) $\nu(\mathbb{P})$ = superhedging error under $\mathbb{P}$,
- (utility maximazation) $\nu(\mathbb{P})$ = maximum utility under $\mathbb{P}$.

These maps are NOT continuous with respect to the weak topology.

The adapted weak topology

How can we define a topology on the space of laws of processes ?

- The adapted weak topology is the smallest topology for which
$$\mathbb{P} \to \inf\{\mathbb{E}^{\mathbb{P}}[\varphi(X_\tau)] : \text{ stopping time } \tau\}$$
is continuous for all continuous, bounded and adapted φ.
- The adapted Wasserstein distance metrizes the adapted weak topology:
$$\mathcal{AW}_1(\mathbb{P},\mathbb{Q}) = \inf\{\mathbb{E}^{\pi}[|X-Y|] : \pi \text{ has marginals } \mathbb{P},\mathbb{Q} \text{ and is } \underline{\text{bicausal}}\}$$
where bicausal means that the transport at each time is independent of future information given the current information.

The adapted weak topology is finer than the weak topology. In particular,
$$\mathcal{W}_1(\mathbb{P},\mathbb{Q}) \le \mathcal{AW}_1(\mathbb{P},\mathbb{Q}).$$
For the aforementioned operators ν, we have
$$|\nu(\mathbb{P}) - \nu(\mathbb{Q})| \lesssim \mathcal{AW}_1(\mathbb{P},\mathbb{Q}).$$

Research question

In many applications, understanding the quantitative behavior of $\mathcal{AW}_1$ is crucial. One natural direction is to establish inequalities of the form
$$\mathcal{AW}_1(\mathbb{P},\mathbb{Q}) \lesssim \mathcal{J}(\mathbb{Q}|\mathbb{P})$$
for certain well-understood functionals $\mathcal{J}$.

- $\mathcal{J}(\mathbb{Q},\mathbb{P}) = \mathrm{TV}(\mathbb{P},\mathbb{Q})$ (TV is the total variation distance): Eckstein–Pammer (2024).
- $\mathcal{J}(\mathbb{Q}|\mathbb{P}) = \sqrt{\mathcal{W}_1(\mathbb{P},\mathbb{Q})}$: Blanchet–Larsson–Park–Wiesel (2024), Acciaio–Hou–Pammer (2025).
- $\mathcal{J}(\mathbb{Q}|\mathbb{P}) = \sqrt{\mathcal{H}(\mathbb{Q}|\mathbb{P})}$ ($\mathcal{H}$ is the relative entropy): Backhoff–Beiglböck–Lin–Zalashko (2017), Beiglböck–Zona (2025).

Question: The inequality $\mathcal{W}_1 \lesssim \sqrt{\mathcal{H}}$ is well-known to characterize the Gaussian concentration. Can we connect the inequality $\mathcal{AW}_1 \lesssim \sqrt{\mathcal{H}}$ with the Gaussian concentration ?

Main result

Theorem (Park, '25). The following are equivalent.

1. $\mathbb{P}$ satisfies the adapted T_1 inequality: for some $C > 0$,
$$\mathcal{AW}_1(\mathbb{P},\mathbb{Q}) \le C\sqrt{\mathcal{H}(\mathbb{Q}|\mathbb{P})} \text{ for all } \mathbb{Q}.$$
2. $\mathbb{P}$ satisfies the T_1 inequality: for some $C > 0$,
$$\mathcal{W}_1(\mathbb{P},\mathbb{Q}) \le C\sqrt{\mathcal{H}(\mathbb{Q}|\mathbb{P})} \text{ for all } \mathbb{Q}.$$
3. $\mathbb{P}$ exhibits the Gaussian concentration: for some $C > 0$,
$$\mathbb{P}(\varphi(X) > \mathbb{E}^{\mathbb{P}}[\varphi(X)] + \varepsilon) \le e^{-C\varepsilon^2} \text{ for all bdd, 1-Lipschitz } \varphi$$
4. $\mathbb{P}$ satisfies the following moment condition: for some $\alpha > 0$,
$$\mathbb{E}^{\mathbb{P}}[\exp(\alpha\,|X|^2)] < \infty.$$

Theorem (Park, '25). If $\mathbb{P}$ satisfies the above statements,
$$\mathcal{AW}_1(\mathbb{P},\mathbb{Q}) \le (2\sqrt{N}+1)\alpha^{-1/2}\left(1 + \log \mathbb{E}^{\mathbb{P}}[\exp(\alpha\,|X|^2)]\right)^{1/2}\sqrt{2\mathcal{H}(\mathbb{Q}|\mathbb{P})}$$
for all $\mathbb{Q}$.

Comparison with existing results

Regularity of kernels

Previous results typically impose certain regularity assumptions on the conditional distributions of $\mathbb{P}$. However, it turns out that such assumptions are not necessary.

Dependence on the temporal dimension N

Our results show that the constant C in $\mathcal{AW}_1 \le C\sqrt{\mathcal{H}}$ grows sublinearly with N, specifically, as $\sqrt{N}$ up to a logarithmic factor. This rate is consistent with the independent case and with the sharp adapted Pinsker inequality, while constants in previous results exhibit exponential growth in N.

Generalization

The theorem extends to any Polish space equipped with an arbitrary metric.

References

[1] Beatrice Acciaio, Songyan Hou, and Gudmund Pammer. Estimating causal distances with non-causal ones. *arXiv preprint arXiv:2506.22421*, 2025.

[2] Julio Backhoff, Mathias Beiglbock, Yiqing Lin, and Anastasiia Zalashko. Causal transport in discrete time and applications. *SIAM Journal on Optimization*, 27(4):2528–2562, 2017.

[3] Mathias Beiglböck and Markus Zona. Pinsker's inequality for adapted total variation. *arXiv preprint arXiv:2506.22106*, 2025.

[4] Jose Blanchet, Martin Larsson, Jonghwa Park, and Johannes Wiesel. Bounding adapted wasserstein metrics. *arXiv preprint arXiv:2407.21492*, 2024.

[5] Stephan Eckstein and Gudmund Pammer. Computational methods for adapted optimal transport. *The Annals of Applied Probability*, 34(1A):675–713, 2024.

[6] Jonghwa Park. On a *t_1* transport inequality for the adapted wasserstein distance. *arXiv preprint arXiv:2507.19215*, 2025.