Sensitivity of robust optimization over an adapted Wasserstein ambiguity set

Yifan Jiang

University of Oxford

September 5, 2022

Table of Contents

Distributionally robust optimization

Framework

$$\inf_{a \in \mathcal{A}} \mathbb{E}^{\mu}[f(X, a)].$$

- Ω is a Polish space.
- $\mathcal{A}$ is an admissible control set.
- f is a cost function.
- μ is the reference distribution (prior knowledge) of the model.

Distributionally robust optimization

Framework

$$V(\delta) = \inf_{a \in \mathcal{A}} \sup_{\nu \in B_\delta(\mu)} \mathbb{E}^\nu[f(X,a)].$$

- Ω is a Polish space.
- $\mathcal{A}$ is an admissible control set.
- f is a cost function.
- μ is the reference distribution (prior knowledge) of the model.
- $B_\delta(\mu)$ is some perturbation of μ with strength δ.

Distributionally robust optimization

Framework

$$V(\delta) = \inf_{a \in \mathcal{A}} \sup_{\nu \in B_\delta(\mu)} \mathbb{E}^\nu[f(X, a)].$$

- Ω is a Polish space.
- $\mathcal{A}$ is an admissible control set.
- f is a cost function.
- μ is the reference distribution (prior knowledge) of the model.
- $B_\delta(\mu)$ is some perturbation of μ with strength δ.

What are we interested in?

The first order approximation of $V(\delta)$ at 0:

$$V'(0) = \lim_{\delta \to 0_+} \frac{V(\delta) - V(0)}{\delta}.$$

Previous results

- KL divergence [Lam, 2016]
  - $\Omega = \mathbf{R}^d$, $\mathcal{A} = \{a\}$, $B_\delta^{KL}(\mu)$ KL ball.
  - $V'(0) = \sqrt{\mathrm{Var}_\mu(f(X,a))}$.
- Wasserstein distance [Bartl et al., 2021]
  - $\Omega = \mathbf{R}^d$, $\mathcal{A}$ convex subset of $\mathbf{R}^n$, $B_\delta^W(\mu)$ Wasserstein-2 ball.
  - $V'(0) = \left(\mathbb{E}^\mu \|\nabla_x f(X,a^*)\|^2\right)^{1/2}$, where a^* is the unique minimizer of $\inf_{a\in\mathcal{A}} \mathbb{E}^\mu[f(X,a)]$.

Previous results

- KL divergence [Lam, 2016]
  - $\Omega = \mathbf{R}^d$, $\mathcal{A} = \{a\}$, $B_\delta^{KL}(\mu)$ KL ball.
  - $V'(0) = \sqrt{\mathrm{Var}_\mu(f(X,a))}$.
- Wasserstein distance [Bartl et al., 2021]
  - $\Omega = \mathbf{R}^d$, $\mathcal{A}$ convex subset of $\mathbf{R}^n$, $B_\delta^W(\mu)$ Wasserstein-2 ball.
  - $V'(0) = \left(\mathbb{E}^\mu \|\nabla_x f(X,a^*)\|^2\right)^{1/2}$, where a^* is the unique minimizer of $\inf_{a \in \mathcal{A}} \mathbb{E}^\mu[f(X,a)]$.

Question

Can we generalize previous results to a dynamic setting?

Table of Contents

What is adapted Wasserstein distance?

Dynamic setting

Let $(\mathbf{R}^d, |\cdot|)$ be the state space, $(\Omega, \mathcal{F}) = ((\mathbf{R}^d)^N, \mathcal{B})$ the canonical space of N-step stochastic processes, X and Y the coordinate processes on $\Omega \times \Omega$.

What is adapted Wasserstein distance?

Dynamic setting

Let $(\mathbf{R}^d, |\cdot|)$ be the state space, $(\Omega, \mathcal{F}) = ((\mathbf{R}^d)^N, \mathcal{B})$ the canonical space of N-step stochastic processes, X and Y the coordinate processes on $\Omega \times \Omega$.

Adapted Wasserstein distance

Give $\mu, \nu \in \mathcal{P}_p(\Omega)$, their adapted Wasserstein 'distance' is defined as

$$\mathcal{AW}_p(\mu, \nu) := \inf_{\pi \in \Pi_c(\mu,\nu)} \Big(\mathbb{E}^\pi \|X - Y\|^p \Big)^{1/p},$$

where $\Pi_c(\mu, \nu)$ is the causal couplings between μ and ν. Metric $\|\cdot\|$ on Ω is given by

$$\|x\|^p := \sum_n |x_n|^p, \quad x = (x_1, \ldots, x_N) \in \Omega = (\mathbf{R}^d)^N.$$

Causal coupling

Definition[1] [Acciaio et al., 2020]

Given $\mu, \nu \in \mathcal{P}_p(\Omega)$, the causal coupling is given by

$$\Pi_c(\mu,\nu) = \{\pi \in \Pi(\mu,\nu) : \theta(\cdot, A) \in {}^{\mu}\mathcal{F}_n^X \quad \text{for any } A \in \mathcal{F}_n^Y\}, \qquad (\star)$$

where θ is the disintegration of π, i.e., $\pi(\mathrm{d}x, \mathrm{d}y) = \theta(x, \mathrm{d}y)\mu(\mathrm{d}x)$.

[1] For equivalent definitions, see [Lassalle, 2018, Backhoff-Veraguas et al., 2020].

Causal coupling

Definition[1] [Acciaio et al., 2020]

Given $\mu, \nu \in \mathcal{P}_p(\Omega)$, the causal coupling is given by

$$\Pi_c(\mu,\nu) = \{\pi \in \Pi(\mu,\nu) : \theta(\cdot, A) \in {}^{\mu}\mathcal{F}_n^X \quad \text{for any } A \in \mathcal{F}_n^Y\}, \qquad (\star)$$

where θ is the disintegration of π, i.e., $\pi(\mathrm{d}x, \mathrm{d}y) = \theta(x, \mathrm{d}y)\mu(\mathrm{d}x)$.

Motivation

Assume a causal coupling $\pi \in \Pi_c(\mu,\nu)$ is generated by a Monge transport map $T : \Omega \to \Omega$, i.e.,

$$\pi = (\mathrm{Id}, T)_{\#}\mu.$$

Then $(\star)$ implies T is adapted (non-anticipative), i.e., for any $A \in \mathcal{F}_n^Y$ we have $T^{-1}(A) \in {}^{\mu}\mathcal{F}_n^X$.

[1] For equivalent definitions, see [Lassalle, 2018, Backhoff-Veraguas et al., 2020].

A classical example

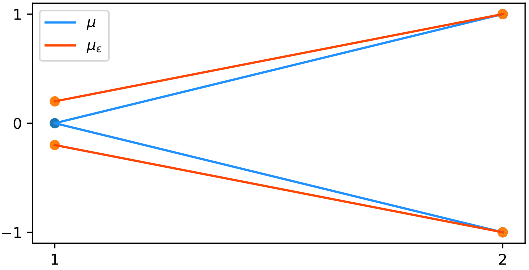
Example

We take $N = 2, d = 1, p = 1$, then $\Omega = \mathbf{R}^2$. Let $\mu = \frac{1}{2}\delta_{(0,1)} + \frac{1}{2}\delta_{(0,-1)}$ and $\mu_\varepsilon = \frac{1}{2}\delta_{(\varepsilon,1)} + \frac{1}{2}\delta_{(-\varepsilon,-1)}$.
Then, as $\varepsilon \to 0$

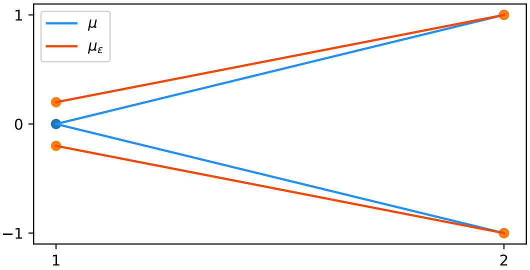
$$\mathcal{W}_1(\mu, \mu_\varepsilon) \to 0,$$

but

$$\mathcal{AW}_1(\mu, \mu_\varepsilon) \to 1!!$$



A classical example



Remark

The optimal stopping problem $\sup_\tau \mathbb{E}[X_\tau]$ shows the remarkable probabilistic difference between μ_ε and μ. In fact, the adapted Wasserstein topology is the coarsest topology on prob measures which makes the optimal stopping problem continuous, see [Backhoff-Veraguas et al., 2020].

Table of Contents

Main results

Theorem[2]

Consider the DRO problem

$$V(\delta) = \inf_{a \in \mathcal{A}} \sup_{\nu \in B_\delta(\mu)} \mathbb{E}^\nu[f(X, a)],$$

where $B_\delta(\mu) = \{\nu : \mathcal{AW}_p(\mu, \nu) \leq \delta\}$. Under suitable assumptions, we have

$$V'(0) = \inf_{a^* \in \mathcal{A}_0^*} \mathbb{E}^\mu \Big[\sum_{n=1}^N \left| \mathbb{E}^\mu \left[\partial_n f(X, a^*) | \mathcal{F}_n \right] \right|^q \Big]^{1/q}.$$

[2]An analogous result for closed-loop control is independently derived in [Bartl and Wiesel, 2022].

Main results

Theorem[2]

Consider the DRO problem

$$V(\delta) = \inf_{a\in\mathcal{A}} \sup_{\nu\in B_\delta(\mu)} \mathbb{E}^\nu[f(X,a)],$$

where $B_\delta(\mu) = \{\nu : \mathcal{AW}_p(\mu,\nu) \le \delta\}$. Under suitable assumptions, we have

$$V'(0) = \inf_{a^*\in\mathcal{A}_0^*} \mathbb{E}^\mu\Big[\sum_{n=1}^N |\mathbb{E}^\mu\big[\partial_n f(X,a^*)|\mathcal{F}_n\big]|^q\Big]^{1/q}.$$

We remark that the sensitivity of DRO problem under the classical Wasserstein perturbation is given by

$$V'(0) = \inf_{a^*\in\mathcal{A}_0^*} \mathbb{E}^\mu\big[\sum_n |\partial_n f(X,a^*)|^q\big]^{1/q}.$$

[2] An analogous result for closed-loop control is independently derived in [Bartl and Wiesel, 2022].

Main results

Assumption

The continuous cost function $f : \Omega \times \mathcal{A} \to \mathbf{R}$ satisfies

- $x \mapsto f(x, a^*)$ is differentiable for any $a^* \in \mathcal{A}^*$. Moreover, $\nabla f(x, a^*)$ is continuous on $\Omega \times \mathcal{A}^*$ and satisfies

$$|\nabla f(x, a^*)| \le C(a^*)(1 + \|x\|^{p-1}),$$

  for any $a^* \in \mathcal{A}^*$, $x \in \Omega$ and some locally bounded function $C : \mathcal{A}^* \to \mathbf{R}$.
- We have $\mathcal{A}^*_\delta \neq \emptyset$ for sufficiently small δ and the limit points of any sequence $\{a^*_k \in \mathcal{A}^*_{\delta_k}\}$ with $\delta_k \to 0$ are contained in $\mathcal{A}^*_0$.

Sketch of the proof (upper bound)

We show the upper bound of $V'(0)$. Let $a^* \in \mathcal{A}_0^*$. Then

$$
\begin{aligned}
V(\delta) - V(0) &\leq \sup_{\nu \in B_\delta(\mu)} \mathbb{E}^\nu[f(X, a^*)] - \mathbb{E}^\mu[f(X, a^*)] \\
&= \sup_{\pi \in C_\delta} \mathbb{E}^\pi[f(Y, a^*) - f(X, a^*)] \\
&= \sup_{\pi \in C_\delta} \int_0^1 \mathbb{E}^\pi[\langle \nabla f(X + \lambda(Y - X), a^*), Y - X \rangle] \, \mathrm{d}\lambda.
\end{aligned}
$$

Let $\mathcal{F}_n^{X,Y} = \mathcal{F}_n^X \vee \mathcal{F}_n^Y$. Then the above inner expectation is equal to

$$
\begin{aligned}
&\sum \mathbb{E}^\pi\left[\mathbb{E}^\pi[\partial_n f(X + \lambda(Y - X), a^*)(Y_n - X_n) | \mathcal{F}_n^{X,Y}]\right] \\
&\leq \sum \|Y_n - X_n\|_{L^p(\pi)} \|\mathbb{E}^\pi[\partial_n f(X + \lambda(Y - X), a^*) | \mathcal{F}_n^{X,Y}]\|_{L^q(\pi)} \\
&\leq \delta \mathbb{E}^\pi\left[\sum |\mathbb{E}^\pi[\partial_n f(X + \lambda(Y - X), a^*) | \mathcal{F}_n^{X,Y}]|^q\right]^{1/q}.
\end{aligned}
$$

Sketch of the proof (upper bound)

Therefore, as $\delta \to 0$

$$
\begin{aligned}
&\limsup_{\delta\to 0} \frac{V(\delta)-V(0)}{\delta} \\
&\leq \limsup_{\delta\to 0} \int_0^1 \mathbb{E}^\pi\Big[\sum_{n=1}^N |\mathbb{E}^\pi\big[\partial_n f(X+\lambda(Y-X),a^*)|\mathcal{F}_n^{X,Y}\big]|^q\Big]^{1/q}\,\mathrm{d}\lambda \\
&\leq \limsup_{\delta\to 0} \int_0^1 \mathbb{E}^\pi\Big[\sum_{n=1}^N |\mathbb{E}^\pi\big[\partial_n f(X,a^*)|\mathcal{F}_n^{X,Y}\big]|^q\Big]^{1/q} \\
&= \mathbb{E}^\mu\Big[\sum_{n=1}^N |\mathbb{E}^\mu\big[\partial_n f(X,a^*)|\mathcal{F}_n\big]|^q\Big]^{1/q}.
\end{aligned}
$$

The last equality comes from the causality of π.

Sketch of the proof (lower bound)

Let $\delta_k \to 0$ and $a_k^* \in \mathcal{A}_{\delta_k}^*$. Therefore, we have

$$\begin{aligned} V(\delta_k) - V(0) &\geq \mathbb{E}^{\nu_k}[f(Y, a_k^*)] - \mathbb{E}^{\mu}[f(X, a_k^*)] \\ &= \mathbb{E}^{\pi_k}[f(Y, a_k^*) - f(X, a_k^*)] \\ &= \int_0^1 \mathbb{E}^{\pi_k}[\langle \nabla f(X + \lambda(Y - X), a_k^*), Y - X \rangle] \, \mathrm{d}\lambda \\ &= \int_0^1 \mathbb{E}^{\mu}[\langle \nabla f(X + \alpha_k \lambda T(X), a_k^*), \alpha_k T(X) \rangle] \, \mathrm{d}\lambda. \end{aligned}$$

Here, the causal coupling $\pi_k \in C_\delta(\mu)$ is given by

$$\pi_k = (\mathrm{Id}, \mathrm{Id} + \alpha_k T)_{\#}\mu,$$

where

$$\begin{aligned} T_n(x_1, \cdots, x_n) \overset{\mu-\text{a.s.}}{:=} & E^{\mu}\left[\partial_n f(X, a^*) | (X_1, \cdots, X_n) = (x_1, \cdots, x_n)\right] \\ & \times \left| \mathbb{E}^{\mu}\left[\partial_n f(X, a^*) | (X_1, \cdots, X_n) = (x_1, \cdots, x_n)\right] \right|^{\frac{q-p}{p}}. \end{aligned}$$

Sketch of the proof (lower bound)

As $\delta \to 0$,

$$\liminf_{\delta\to 0} \frac{V(\delta) - V(0)}{\delta}$$
$$\geq \lim_{k\to\infty} \int_0^1 \mathbb{E}^\mu[\langle \nabla f(X + \alpha_k \lambda T(X), a_k^*), T(X)\rangle] \, \mathrm{d}\lambda$$
$$= \mathbb{E}^\mu[\langle \nabla f(X, a^*), T(X)\rangle]$$
$$= \mathbb{E}^\mu\Big[\sum_{n=1}^N \big|\mathbb{E}^\mu\big[\partial_n f(X, a^*) \big| \mathcal{F}_n\big]\big|^q\Big].$$

Table of Contents

Martingale constraint

Theorem

Let $\mathcal{M}_p$ *be the space of martingale probability measures on* $(\Omega, \mathcal{F})$ *with finite* p*-moment. We consider*

$$V(\delta) = \inf_{a \in \mathcal{A}} \sup_{\nu \in M_\delta(\mu)} \mathbb{E}^\nu[f(X,a)],$$

where $M_\delta(\mu) = B_\delta(\mu) \cap \mathcal{M}_p$*. Under suitable assumptions, we have*

$$V'(0) = \inf_{a^* \in \mathcal{A}_0^*} \inf_{\Lambda} \mathbb{E}^\mu \left[\sum_{n=1}^{N} \left| \mathbb{E}^\mu \left[\partial_n f(X, a^*) | \mathcal{F}_n \right] + \Lambda_{n-1}(X) - \Lambda_n(X) \right|^q \right]^{1/q},$$

where Λ *is non-anticipative and* $\Lambda_n : \Omega \to \mathbf{R}^d$ *is smooth and compactly supported.*

Martingale constraint

Asian option

We consider a discrete monitoring Asian call with strike K. Let μ be the risk neutral measure and set the risk-free interest as 0. Then the price of the Asian call is given by $\mathbb{E}^\mu[f(X)]$, where f is the payoff

$$f(x) = \Big(\sum_{n=1}^{N} x_n - K\Big)^+.$$

Let $p = 2$, then the sensitivity of the price $V'(0)$ is explicitly given by

$$\begin{aligned} V'(0) &= \inf_{\Lambda} \mathbb{E}^\mu \Big[\sum_{n=1}^{N} \big|\mathbb{E}^\mu\big[\partial_n f(X)|\mathcal{F}_n\big] + \Lambda_{n-1}(X) - \Lambda_n(X)\big|^2\Big]^{1/2} \\ &= \mathbb{E}^\mu \Big[\sum_{n=1}^{N} \mu\big(\sum_{n=1}^{N} X_n \geq K|\mathcal{F}_n\big)^2\Big]^{1/2}. \end{aligned}$$

Optimization with cost of weak-type

For the simplicity, we restrict ourselves to $p = 2$ and two-step processes. Consider a problem depending on the condition law of the reference distribution, e.g.

$$V(\delta) = \sup_{\nu \in \tilde{B}_\delta(\mu)} \mathbb{E}^\nu[f(X_1, \mathrm{Law}(X_2|X_1))].$$

Optimization with cost of weak-type

For the simplicity, we restrict ourselves to $p = 2$ and two-step processes. Consider a problem depending on the condition law of the reference distribution, e.g.

$$V(\delta) = \sup_{\nu \in \tilde{B}_\delta(\mu)} \mathbb{E}^\nu[f(X_1, \nu^{X_1})],$$

where ν^{x_1} is the disintegration of ν with respect to its first marginal.

Optimization with cost of weak-type

For the simplicity, we restrict ourselves to $p = 2$ and two-step processes. Consider a problem depending on the condition law of the reference distribution, e.g.

$$V(\delta) = \sup_{\nu \in \tilde{B}_\delta(\mu)} \mathbb{E}^\nu[f(X_1, \nu^{X_1})],$$

where ν^{x_1} is the disintegration of ν with respect to its first marginal.

Lions' derivative

Given a function $\phi : \mathcal{P}_2(\mathbf{R}^d) \to \mathbf{R}$, the lift of ϕ on some probability space $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$ is given by

$$\Phi(X) = \phi(\mathrm{Law}(X)),$$

for any $X \in L^2(\hat{\Omega}, \hat{\mathbb{P}})$. The Lions' derivative of ϕ at μ is defined as

$$\partial\phi(\mu)(X) := \mathscr{D}\Phi(X),$$

where $\mathrm{Law}(X) = \mu$ and $\mathscr{D}\Phi$ is the Fréchet derivative of Φ.

Optimization with cost of weak-type

Theorem

We consider

$$V(\delta) = \sup_{\nu \in \tilde{B}_\delta(\mu)} \mathbb{E}^\nu[f(X_1, \nu^{X_1})].$$

Under suitable assumptions, $V'(0)$ is given by

$$V'(0) = \mathbb{E}^{\mu_1}\left[|\partial_1 f(X_1, \mu^{X_1})|^2 + \hat{\mathbb{E}}[|\partial_2 f(X_1, \mu^{X_1})(\hat{X}^{X_1})|^2]\right]^{1/2},$$

where $\mu(\mathrm{d}x_1, \mathrm{d}x_2) = \mu_1(\mathrm{d}x_1)\mu^{x_1}(\mathrm{d}x_2)$ and $\hat{X}^{x_1}$ is a random variable on $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$ following μ^{x_1}.

Optimization with cost of weak-type

Optimal stopping problem

We consider a simple two-step optimal stopping problem

$$V(0) = \sup_{\tau} \mathbb{E}^{\mu}[g(X_\tau)].$$

By Snell envelope, the value function has the form of

$$\mathbb{E}^{\mu}[f(X_1, \mu^{X_1})] := \mathbb{E}^{\mu}[\max(g(X_1), \mathbb{E}^{\mu}[g(X_2)|X_1])].$$

Under suitable conditions, the sensitivity of the optimal stopping problem is explicitly given by

$$V'(0) = \mathbb{E}^{\mu_1}\left[|g'(X_1)|^2 \mathbb{1}_A + \hat{\mathbb{E}}|g'(\hat{X}^{X_1})|^2 \mathbb{1}_{A^c}\right]^{1/2},$$

where $A = \{g(X_1) > \mathbb{E}^{\mu}[g(X_2)|X_1]\}$ and $\hat{X}^{x_1}$ is a random variable on $(\hat{\Omega}, \hat{\mathcal{F}}, \hat{\mathbb{P}})$ following μ^{x_1}.

Main references I

[Acciaio et al., 2020] Acciaio, B., Backhoff-Veraguas, J., and Zalashko, A. (2020).
Causal optimal transport and its links to enlargement of filtrations and continuous-time stochastic optimization.
Stochastic Processes and their Applications, 130(5):2918–2953.

[Backhoff-Veraguas et al., 2020] Backhoff-Veraguas, J., Bartl, D., Beiglböck, M., and Eder, M. (2020).
All adapted topologies are equal.
Probab. Theory Relat. Fields, 178(3-4):1125–1172.

[Bartl et al., 2021] Bartl, D., Drapeau, S., Obłój, J., and Wiesel, J. (2021).
Sensitivity analysis of Wasserstein distributionally robust optimization problems.
Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 477(2256):20210176.

Main references II

[Bartl and Wiesel, 2022] Bartl, D. and Wiesel, J. (2022).
Sensitivity of multi-period optimization problems in adapted Wasserstein distance.

[Lam, 2016] Lam, H. (2016).
Robust Sensitivity Analysis for Stochastic Systems.
Mathematics of OR, 41(4):1248–1275.

[Lassalle, 2018] Lassalle, R. (2018).
Causal transport plans and their Monge–Kantorovich problems.
Stochastic Analysis and Applications, 36(3):452–484.

Thank you!