# Circumventing Compactness in Conditional-Moment Optimal Transport Duality

## A Perturbation Approach

Louis L. Chen[1] **Jake Roth**[2] Johannes O. Royset[3]

March 21, 2026

[1]Operations Research, Naval Postgraduate School.
[2]Industrial and Systems Engineering, University of Minnesota.
[3]Industrial and Systems Engineering, University of Southern California.

# Recent work in DRO

DRO problem for uncertain $U$ taking values in $\mathcal{U}$:

$$\underset{x\in\mathcal{X}}{\text{infimum}} \qquad \mathrm{UQ}(x) \coloneqq \sup_{\mu\in\mathcal{A}(\hat{\mu})} \mathbb{E}_\mu[f(x,U)] \tag{DRO}$$

Popular ambiguity sets / models:

**Wasserstein / OT [MK18; BM19; ZYG24; GK23; ASD26]**

$$\sup_{\gamma\in\mathcal{P}(\mathcal{U}\times\mathcal{U})} \mathbb{E}_\gamma[f(U)] \ : \ \mathbb{E}_\gamma[d(\hat{U},U)] \le \rho, \quad \pi_1\gamma = \hat{\mu}$$

"influence how *values* behave"

# Recent work in DRO

DRO problem for uncertain $U$ taking values in $\mathcal{U}$:

$$\underset{x\in\mathcal{X}}{\text{infimum}} \qquad \mathrm{UQ}(x) := \sup_{\mu\in\mathcal{A}(\hat\mu)} \mathbb{E}_\mu[f(x,U)] \tag{DRO}$$

Popular ambiguity sets / models:

**$\phi$-divergence [BT07; BL15; AH21]**

$$\sup_{\mu\in\mathcal{P}(\mathcal{U})} \mathbb{E}_\mu[f(U)] \ : \ \mathbb{D}_\phi(\hat\mu,\mu) \le \rho$$

"influence how *weights* behave"

# Recent work in DRO

DRO problem for uncertain $U$ taking values in $\mathcal{U}$:

$$\operatorname*{infimum}_{x\in\mathcal{X}} \qquad \mathrm{UQ}(x) \coloneqq \sup_{\mu\in\mathcal{A}(\hat{\mu})} \mathbb{E}_\mu[f(x,U)] \tag{DRO}$$

Popular ambiguity sets / models:

**Sinkhorn [WGX25; WM23; Dap+23; YZL25]**

$$\sup_{\gamma\in\mathcal{P}(\mathcal{U}\times\mathcal{U})} \mathbb{E}_\gamma[f(U)] \ : \ \mathbb{E}_\gamma[d(\hat{U},U)] + \beta\,\mathbb{D}(\gamma,\hat{\eta}\otimes\eta) \leq \rho, \quad \pi_1\gamma = \hat{\mu}$$

"interpolate between OT and $\phi$-divergence"

# Recent work in DRO

DRO problem for uncertain $U$ taking values in $\mathcal{U}$:

$$\underset{x \in \mathcal{X}}{\text{infimum}} \qquad \mathrm{UQ}(x) \coloneqq \sup_{\mu \in \mathcal{A}(\hat{\mu})} \mathbb{E}_\mu[f(x, U)] \tag{DRO}$$

Popular ambiguity sets / models:

**Unified model: CM-OT [Bla+25] handles each**

$$\sup_{\mu \in \mathcal{P}(\mathcal{U})} \mathbb{E}_\mu[f(U)] \ : \ \mathbb{M}(\hat{\mu}, \mu) \leq \rho$$

"influence how *values* and *weights* behave"

# CM-OT discrepancy $\mathbb{M}(\hat{\mu}, \mu)$ [Bla+25]

- Space: $\mathcal{U} = \mathcal{V} \times \mathcal{W}$ of values and weights;
- Objective $f : \mathcal{U} \to \overline{\mathbb{R}}$ (dropping dependence on "$x$");
- CM-OT discrepancy

$$\mathbb{M}_h(\hat{\mu}, \mu) := \left\{ \begin{array}{ll} \underset{\gamma \in \Gamma(\hat{\mu}, \mu)}{\text{infimum}} & \mathbb{E}_\gamma[c(\hat{V}, \hat{W}; V, W)] \\ \text{s. t.} & A\gamma = 0 \in L^1, \ \hat{\nu}\text{-a.s.} \end{array} \right\}$$

- Primal UQ

$$(\mathsf{P}) \ : \ \sup_{\mu \in \mathcal{P}(\mathcal{U})} \mathbb{E}_\mu[f] : \mathbb{M}(\hat{\mu}, \mu) \leq \rho$$

# CM-OT duality

**Upper bounds for** (P):

$$(\mathsf{P}) = \sup_{\mu \in \mathcal{P}(\mathcal{U})} \left\{ \mathbb{E}_\mu[f] \ : \ \mathbb{M}(\hat{\mu}, \mu) \leq \rho \right\}$$

# CM-OT duality

**Upper bounds for** (P):

$$(\mathsf{P}) = \sup_{\mu \in \mathcal{P}(\mathcal{U})} \Big\{ \mathbb{E}_{\mu}[f] \ : \ \inf_{\gamma \in \Gamma(\hat{\mu}, \mu)} \sup_{\psi \in \Psi} \big\{ \mathbb{E}_{\gamma}[c] + \langle \psi, A\gamma \rangle \big\} \leq \rho \Big\}$$

# CM-OT duality

**Upper bounds for** (P):

$$(\mathsf{P}) \leq \sup_{\mu \in \mathcal{P}(\mathcal{U})} \Big\{ \mathbb{E}_\mu[f] \ : \ \sup_{\psi \in \Psi} \inf_{\gamma \in \Gamma(\hat{\mu}, \mu)} \big\{ \mathbb{E}_\gamma[c] + \langle \psi, A\gamma \rangle \big\} \leq \rho \Big\}$$

# CM-OT duality

**Upper bounds for** (P):

$$(\mathsf{P}) \le \sup_{\mu\in\mathcal{P}(\mathcal{U})}\Big\{\inf_{\lambda\ge 0}\Big\{\mathbb{E}_\mu[f]+\lambda\big(\rho-\sup_{\psi\in\Psi}\inf_{\gamma\in\Gamma(\hat\mu,\mu)}\{\mathbb{E}_\gamma[c]+\langle\psi,A\gamma\rangle\}\big)\Big\}\Big\}$$

# CM-OT duality

**Upper bounds for** (P):

$$(\mathsf{P}) \leq \inf_{\lambda \geq 0} \sup_{\mu \in \mathcal{P}(\mathcal{U})} \Big\{ \mathbb{E}_{\mu}[f] + \lambda\big(\rho - \sup_{\psi \in \Psi} \inf_{\gamma \in \Gamma(\hat{\mu}, \mu)} \{ \mathbb{E}_{\gamma}[c] + \langle \psi, A\gamma \rangle \}\big) \Big\}$$

# CM-OT duality

**Upper bounds for** (P):

$$(\mathsf{P}) \le \inf_{\substack{\lambda \ge 0 \\ \psi \in \Psi}} \lambda\rho + \sup_{\gamma \in \Gamma_{\hat{\mu}}} \mathbb{E}_\gamma\Big[f(U) - \psi(\hat{V}) \cdot (W - h(\hat{V})) - \lambda c(\hat{U}; U)\Big] \;:\; (\mathsf{D}_\Psi)$$

# CM-OT duality

**Upper bounds for** (P):

$$
\begin{aligned}
(\mathsf{P}) &\le \inf_{\substack{\lambda\ge 0\\ \psi\in\Psi}} \lambda\rho + \sup_{\gamma\in\Gamma_{\hat\mu}} \mathbb{E}_\gamma\Big[f(U) - \psi(\hat V)\cdot(W - h(\hat V)) - \lambda c(\hat U; U)\Big] \;:\; (\mathsf{D}_\Psi)\\
&\le \inf_{\substack{\lambda\ge 0\\ \psi\in\Psi}} \lambda\rho + \mathbb{E}_{\hat\mu}\Big[\sup_{u\in\mathcal{U}} f(u) - \psi(\hat V)\cdot(w - h(\hat V)) - \lambda c(\hat U; u)\Big] \;:\; (\mathsf{D}_\Psi + \mathsf{IP})
\end{aligned}
$$

# CM-OT duality

**Upper bounds for** (P):

$$(\mathsf{P}) \leq \inf_{\substack{\lambda \geq 0 \\ \psi \in \Psi}} \lambda\rho + \sup_{\gamma \in \Gamma_{\hat{\mu}}} \mathbb{E}_\gamma \Big[ f(U) - \psi(\hat{V}) \cdot (W - h(\hat{V})) - \lambda c(\hat{U}; U) \Big] \;:\; (\mathsf{D}_\Psi)$$

$$\leq \inf_{\substack{\lambda \geq 0 \\ \psi \in \Psi}} \lambda\rho + \mathbb{E}_{\hat{\mu}} \Big[ \sup_{u \in \mathcal{U}} f(u) - \psi(\hat{V}) \cdot (w - h(\hat{V})) - \lambda c(\hat{U}; u) \Big] \;:\; (\mathsf{D}_\Psi + \mathsf{IP})$$

**CM-OT duality** $[(\mathsf{P}) = (\mathsf{D}_\Psi + \mathsf{IP})]$:

# CM-OT duality

**Upper bounds for** (P):

$$(\mathsf{P}) \le \inf_{\substack{\lambda\ge 0\\ \psi\in\Psi}} \lambda\rho + \sup_{\gamma\in\Gamma_{\hat\mu}} \mathbb{E}_\gamma\Big[f(U) - \psi(\hat V)\cdot(W - h(\hat V)) - \lambda c(\hat U;U)\Big] \; : \; (\mathsf{D}_\Psi)$$

$$\le \inf_{\substack{\lambda\ge 0\\ \psi\in\Psi}} \lambda\rho + \mathbb{E}_{\hat\mu}\Big[\sup_{u\in\mathcal{U}} f(u) - \psi(\hat V)\cdot(w - h(\hat V)) - \lambda c(\hat U;u)\Big] \; : \; (\mathsf{D}_\Psi + \mathsf{IP})$$

**CM-OT duality** $[(\mathsf{P}) = (\mathsf{D}_\Psi + \mathsf{IP})]$:

I. $[(\mathsf{P}) = (\mathsf{D}_\Psi)]$ zero duality gap; and

# CM-OT duality

**Upper bounds for** (P):

$$(\mathsf{P}) \leq \inf_{\substack{\lambda\geq 0\\ \psi\in\Psi}} \lambda\rho + \sup_{\gamma\in\Gamma_{\hat{\mu}}} \mathbb{E}_\gamma\Big[f(U) - \psi(\hat{V})\cdot(W - h(\hat{V})) - \lambda c(\hat{U};U)\Big] \;:\; (\mathsf{D}_\Psi)$$

$$\leq \inf_{\substack{\lambda\geq 0\\ \psi\in\Psi}} \lambda\rho + \mathbb{E}_{\hat{\mu}}\Big[\sup_{u\in\mathcal{U}} f(u) - \psi(\hat{V})\cdot(w - h(\hat{V})) - \lambda c(\hat{U};u)\Big] \;:\; (\mathsf{D}_\Psi + \mathsf{IP})$$

**CM-OT duality** $[(\mathsf{P}) = (\mathsf{D}_\Psi + \mathsf{IP})]$:

I. $[(\mathsf{P}) = (\mathsf{D}_\Psi)]$ zero duality gap; and
II. $[(\mathsf{D}_\Psi) = (\mathsf{D}_\Psi + \mathsf{IP})]$ Interchangeability Principle (IP) [Roc74; ZYG24].

# When $\mathcal{U}$ is compact

Space of multipliers $\Psi = \mathcal{C}_b(\mathcal{V})$: Dual to compactly supported measures

Zero duality gap: Sion minimax for interchanging $\sup_{\gamma \in \Gamma_{\hat{\mu}}} \inf_{\lambda \geq 0,\, \psi \in \Psi}$

IP: Ensures existence of a solution map for IP

**when $\mathcal{U}$ is not compact...** **(e.g., $\hat{\mu}$ is Gaussian)**

**CM-OT duality fails in general:** $[(\mathsf{P}) \neq (\mathsf{D}_\Psi)]$ **for unbounded** $f$

# Our results: when $\mathcal{U}$ is not compact

**Theorem 1 (Zero gap: $(\mathsf{P}) = (\mathsf{D}_{L^\infty})$)**

*If $f$ is bounded above; $c$ is bounded below; and two assumptions (*pointwise correction *and* strict feasibility*) hold; then* $(\mathsf{P}) = (\mathsf{D}_{L^\infty})$.

**Proof idea:** use *characterization* of "zero gap" in perturbation duality

**Theorem 2 (IP: $(\mathsf{D}_{L^\infty}) = (\mathsf{D}_{\mathcal{C}_b} + \mathsf{IP})$)**

*If in addition: $f$ is usc; $c$ is lsc; and $\mathcal{V}$ is Polish and $\mathcal{W}$ is bounded; then* $(\mathsf{D}_{L^\infty}) = (\mathsf{D}_{\mathcal{C}_b} + \mathsf{IP})$.

**Proof idea:** Lusin's theorem + Tietze extension theorem

# Perturbation approach when $\mathcal{U}$ is not compact

**Look for perturbation assumptions:**

1. What if we have a strictly feasible measure $\mu^0$?

   Can be helpful for existence of $\lambda$ for perturbed budget constraints $\mathbb{M}(\hat{\mu}, \mu) \leq \rho + \tau$

   Not enough...

2. What if we could perturb the CM constraint rhs $h$ up/down?

   Could we "correct" any failures of the CM constraint?

# Perturbation argument I: closedness goal

**Starting point**:

$$(\mathsf{P}) \geq (\mathsf{P}') \quad \left\{ \begin{array}{ll} \sup\limits_{\gamma \in \Gamma_{\hat{\mu}}} & \mathbb{E}_\gamma[f] \\ \text{s. t.} & \mathbb{E}_\gamma[c] \leq \rho, \\ & \mathbb{E}_\gamma[W \mid \hat{V}] - h(\hat{V}) = 0, \ \hat{\nu}\text{-a.s.} \end{array} \right\}$$

# Perturbation argument I: closedness goal

**Starting point**:

$$p(\tau,\theta)=\left\{\begin{array}{ll}\sup\limits_{\gamma\in\Gamma_{\hat\mu}} & \mathbb{E}_\gamma[f] \\ \text{s. t.} & \mathbb{E}_\gamma[c]\le\rho+\tau, \\ & \mathbb{E}_\gamma[W\mid\hat V]-h(\hat V)=\theta,\ \hat\nu\text{-a.s.}\end{array}\right\}$$

**Goal**: Show [Roc70; Pon80; ET99; Zal02]

$$(\mathsf{D})\equiv(\operatorname{cl}p)(0,0)\equiv\limsup_{(\tau,\theta)\to 0}p(\tau,\theta)\le p(0,0)\equiv(\mathsf{P}). \qquad (*)$$

# Perturbation argument II: sequences

**Main question**: How do nearby solutions to nearby problems $p(\tau, \theta)$ relate to the unperturbed problem $p(0, 0)$?

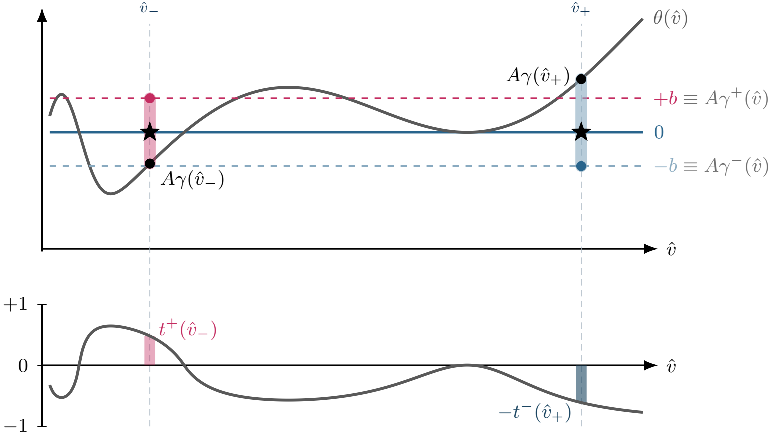
$$\limsup_{(\tau,\theta)\to 0} p(\tau, \theta) \leq p(0, 0) \qquad (*)$$

Step 1: there exists $\{\tau_r, \theta_r\}_{r=1}^{\infty} \subseteq \mathbb{R} \times L^1$ s.t. $(\tau_r, \theta_r) \to 0$ $p(\tau_r, \theta_r) > -\infty$

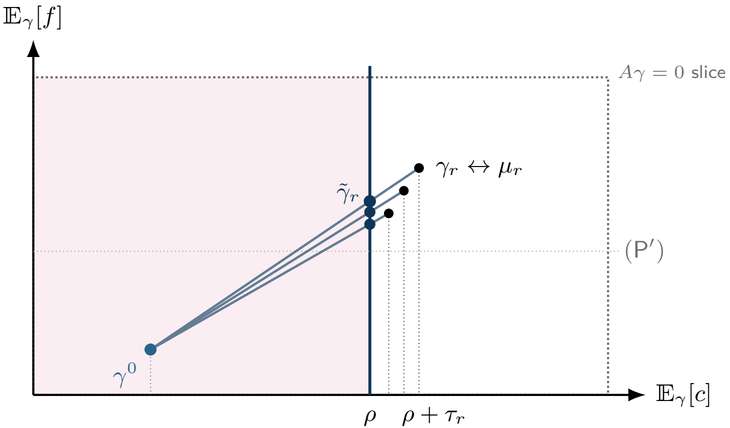
Step 2: leverage *pointwise correction*

Step 3: leverage *strict feasibility*

# Assumption: pointwise correction



# Assumption: strict feasibility




$$\limsup_{(\tau,\theta)\to 0} p(\tau, \theta) \leq p(0,0)$$

# Discussion

- Assumptions in earlier works *implied* that $f, c$ were bounded over $\mathcal{W} \times \mathcal{V}$ and $\mathcal{W}$;
- *Strict feasibility* holds if $\rho$ is replaced with $\rho + \epsilon$ for any $\epsilon > 0$;
- *Pointwise correction* holds if $c, f$ are Lipschitz continuous;
- In Theorems 1 and 2, neither primal nor dual attainment is guaranteed; see arXiv (uploaded and processing!) for counterexamples.

# Questions? Thanks!

**Draft:** arXiv:2603.20503

# References I


[AH21] Rohit Agrawal and Thibaut Horel. "Optimal Bounds between $f$-Divergences and Integral Probability Metrics." In: *J. of Mach. Learn. Res.* 22 (2021).

[ASD26] Liviu Aolaritei, Soroosh Shafiee, and Florian Dörfler. "Wasserstein distributionally robust estimation in high dimensions: Performance analysis and optimal hyperparameter tuning." In: *Math. Prog.* (2026).

[BL15] Güzin Bayraksan and David K. Love. "Data-Driven Stochastic Programming Using $\phi$-Divergences." In: *INFORMS TutORials in Oper. Res.* (2015).

[Bla+25] J. Blanchet et al. *Unifying Distributionally Robust Optimization via Optimal Transport Theory*. arXiv:2308.05414. 2025.

[BM19] José Blanchet and Karthyek Murthy. "Quantifying Distributional Model Risk Via Optimal Transport." In: *Math. Oper. Res.* 44.2 (2019).

# References II


[BT07] Aharon Ben-Tal and Marc Teboulle. "An Old-New Concept of Convex Risk Measures: The Optimized Certainty Equivalent." In: *Mathematical Finance* 17.3 (2007), pp. 449–476.

[Dap+23] Charles Dapogny et al. "Entropy-regularized Wasserstein distributionally robust shape and topology optimization." In: *Structural and Multidisciplinary Optim.* 66.42 (2023).

[ET99] I. Ekeland and R. Témam. *Convex Analysis and Variational Problems*. SIAM, 1999.

[GK23] R. Gao and A. Kleywegt. "Distributionally Robust Stochastic Optimization with Wasserstein Distance." In: *Math. Oper. Res.* 48.2 (2023).

[MK18] Peyman Mohajerin Esfahani and Daniel Kuhn. "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations." In: *Math. Prog.* 171.1–2 (2018).

[Pon80] J. Ponstein. *Approaches to the Theory of Optimization*. Cambridge University Press, 1980.

# References III

[Roc70] R. T. Rockafellar. *Convex Analysis*. Vol. 11. Princeton University Press, 1970.

[Roc74] R T Rockafellar. *Conjugate Duality and Optimization*. Conference Board of Math. Sciences Series, SIAM Publications, 1974.

[WGX25] Jie Wang, Rui Gao, and Yao Xie. "Sinkhorn Distributionally Robust Optimization." In: *Oper. Res.* (Oct. 2025).

[WM23] Franck Iutzeler Waïss Azizian and Jérôme Malick. "Regularization for Wasserstein distributionally robust optimization." In: *ESAIM: Control, Optimisation and Calculus of Variations* 29.33 (2023).

[YZL25] Yufeng Yang, Yi Zhou, and Zhaosong Lu. *Nested Stochastic Algorithm for Generalized Sinkhorn distance-Regularized Distributionally Robust Optimization*. 2503.22923. 2025.

[Zal02] C. Zalinescu. *Convex Analysis in General Vector Spaces*. World Scientific, 2002.

# References IV

[ZYG24] L. Zhang, J. Yang, and R. Gao. "A Short and General Duality Proof for Wasserstein Distributionally Robust Optimization." In: *Oper. Res.* 73.4 (2024).

# Appendices

# When does *pointwise correction* hold?

Suppose:

- *strict feasibility* holds: $\gamma^0 : \mathbb{E}_{\gamma^0}[c] < \rho$ and $A\gamma^0 = 0$;
- $h(\hat{V}) \pm \beta \in \mathcal{W}$
- $c$ is $L$-Lipschitz in $(\hat{v}, \hat{w}; v, \cdot)$

Then:

- $\gamma^0$ is nearly good enough
- if $f$ are continuous in the $w$-component, then we can trade-of slack in the transport budget to perturb $\gamma^0$'s $w$-component up/down slightly to $h \pm b$
- and if $f > -\infty$

# Unbounded $f$

Suppose: $\mathcal{V} := \mathbb{R}$ $\mathcal{W} := \mathbb{R}_+$ $f(v,w) := v \cdot (w-1)$, $h(\hat{V}) \equiv 1$, $\rho > 0$, $c(\hat{v},\hat{w};v,w)$ take value $+\infty$ if $v \neq \hat{v}$ and zero otherwise.

Also let $\hat{\nu}$ be any probability distribution on $\mathcal{V}$ that is not essentially bounded above, i.e., $\hat{\nu}(\hat{V} > t) > 0$ for every $t \in \mathbb{R}$, and let $\hat{\mu} := \hat{\nu} \otimes \delta_1$ be the product measure of $\hat{\nu}$ with the point-mass on 1.

Finally let $\Psi := L^\infty(\hat{\nu})$.

Then $0 = (\mathsf{P}) < (\mathsf{D}_\Psi) = +\infty$.