Oracle-Based Distributionally Robust Optimization under Optimal Transport Ambiguity Sets

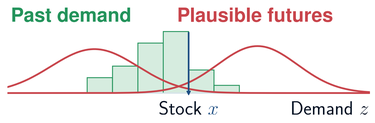
Guixian Chen[1] Salar Fattahi[1] Soroosh Shafiee[2]
[1] University of Michigan [2] Cornell University

INDUSTRIAL & OPERATIONS ENGINEERING
UNIVERSITY OF MICHIGAN

arXiv

Motivation

Limited data and changing conditions make future outcomes uncertain.



Inventory example

Too little stock: shortages.
Too much stock: unsold inventory.

DRO minimizes the worst expected loss over plausible distributions.

Challenge: conic reformulations introduce many auxiliary variables.
Typical interior-point solvers scale poorly with problem size.

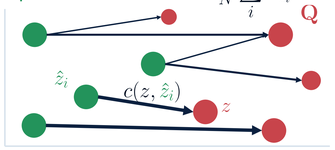
Problem setup

$$\mathrm{OT}_c(\mathbf{Q}, \widehat{\mathbf{P}}) := \min_{\pi \in \Gamma(\mathbf{Q}, \widehat{\mathbf{P}})} \mathbb{E}_{(z,\hat{z})\sim\pi}[c(z, \hat{z})]$$

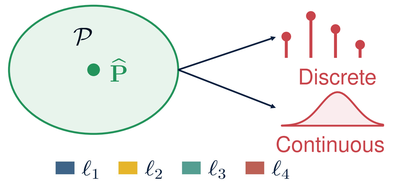
$\Gamma(\mathbf{Q}, \widehat{\mathbf{P}})$: couplings with marginals $\mathbf{Q}$ and $\widehat{\mathbf{P}}$, respectively.

Optimal transport

Empirical center: $\widehat{\mathbf{P}} = \frac{1}{N}\sum_i \delta_{\hat{z}_i}$



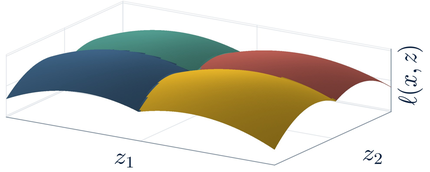
An OT ball of distributions



Piecewise Structure:

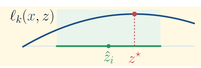
$\ell(x, z) = \max_{k\in[K]} \ell_k(x, z)$

ℓ_k is convex in $x \in \mathcal{X} \subseteq \mathbb{R}^n$
and concave in $z \in \mathcal{Z} \subseteq \mathbb{R}^m$



Oracle-based methods

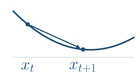
Local Oracle: $F_{ik}(u) := \max_{z\in\mathcal{Z}}\{\ell_k(x, z) : c(z, \hat{z}_i) \le u\}$



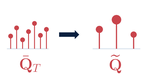
- **Worst-case expectation: budget allocation**
  $\sup_{\mathbf{Q}\in\mathcal{P}} \mathbb{E}_{z\sim\mathbf{Q}}[\ell(x, z)]$, where $\mathcal{P} := \{\mathbf{Q} \in \mathcal{P}(\mathcal{Z}) : \mathrm{OT}_c(\mathbf{Q}, \widehat{\mathbf{P}}) \le \rho\}$



- **Primal DRO: online convex optimization**
  $\min_{x\in\mathcal{X}} \sup_{\mathbf{Q}\in\mathcal{P}} \mathbb{E}_{z\sim\mathbf{Q}}[\ell(x, z)]$



- **Dual DRO: primal DRO, then tangent compression**
  $\sup_{\mathbf{Q}\in\mathcal{P}} \min_{x\in\mathcal{X}} \mathbb{E}_{z\sim\mathbf{Q}}[\ell(x, z)]$



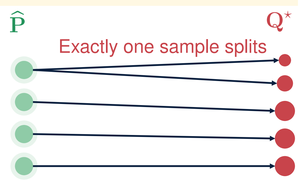
Main reduction: minimal support

Fix x. Reformulate over weights α and shifts v:

$$\sup_{\mathbf{Q}\in\mathcal{P}} \mathbb{E}_{z\sim\mathbf{Q}}[\ell(x, z)] = \max_{\alpha\ge 0,\, v} \frac{1}{N}\sum_{i,k} \alpha_{ik}\ell_k(x, \hat{z}_i + v_{ik})$$
$$\text{s.t.} \quad \sum_k \alpha_{ik} = 1, \quad \frac{1}{N}\sum_{i,k} \alpha_{ik} c(\hat{z}_i + v_{ik}, \hat{z}_i) \le \rho, \quad \hat{z}_i + v_{ik} \in \mathcal{Z}$$

- Fix optimal shifts $v^\star$: LP in weights α.
- Take an optimal basic feasible solution.

N mass constraints + 1 budget constraint
➡ $|\mathrm{supp}(\mathbf{Q}^\star)| \le N + 1$



Worst-case expectation

Allocate the shared transport budget across samples:

$$\sup_{\mathbf{Q}\in\mathcal{P}} \mathbb{E}_{z\sim\mathbf{Q}}[\ell(x, z)] = \max_{b\ge 0} \frac{1}{N}\sum_i V_i(b_i) \quad \text{s.t.} \quad \frac{1}{N}\sum_i b_i \le \rho$$

$$V_i(b) := \max_{\substack{\alpha\ge 0,\, v \\ \hat{z}_i + v_k \in \mathcal{Z}}} \sum_k \alpha_k \ell_k(x, \hat{z}_i + v_k) \quad \text{s.t.} \quad \sum_k \alpha_k = 1,\ \sum_k \alpha_k c(\hat{z}_i + v_k, \hat{z}_i) \le b$$

$$b_i(\lambda) \in \arg\max_{b\ge 0}\{V_i(b) - \lambda b\}, \quad i \in [N]$$

Dualize budget: bisection in $\lambda \ge 0$

Sample 1	Sample i	Sample N
Golden-section in b_1	Golden-section in b_i	Golden-section in b_N
All loss-piece pairs	All loss-piece pairs	All loss-piece pairs
$(1,2) \cdots (K-1, K)$	$(1,2) \cdots (K-1, K)$	$(1,2) \cdots (K-1, K)$
Nested golden-section over (α, β)	Nested golden-section over (α, β)	Nested golden-section over (α, β)

$$V_i(b) = \max_{1\le k_1<k_2\le K} V_i^{(k_1,k_2)}(b) \qquad \text{number of pairs} = \binom{K}{2}$$

$$V_i^{(k_1,k_2)}(b) = \max_{\alpha\in[0,1],\,\beta\in[0,b]} \left\{\alpha F_{ik_1}\left(\frac{\beta}{\alpha}\right) + (1-\alpha) F_{ik_2}\left(\frac{b-\beta}{1-\alpha}\right)\right\}$$

Primal DRO: distributional best-response

$f(x, \mathbf{Q}) := \mathbb{E}_{z\sim\mathbf{Q}}[\ell(x, z)]$

Worst-case response

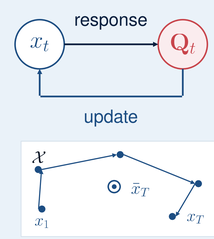
$$\mathbf{Q}_t \approx \arg\max_{\mathbf{Q}\in\mathcal{P}} f(x_t, \mathbf{Q})$$

Decision update

$$x_{t+1} = \Pi_{\mathcal{X}}(x_t - \eta_t g_t), \qquad g_t \in \partial_x f(x_t, \mathbf{Q}_t)$$

Averaging

$$\bar{x}_T = \frac{1}{T}\sum_{t=1}^T x_t \qquad \bar{\mathbf{Q}}_T = \frac{1}{T}\sum_{t=1}^T \mathbf{Q}_t$$

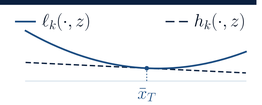


$$\mathrm{Gap}(\bar{x}_T, \bar{\mathbf{Q}}_T) := \sup_{\mathbf{Q}\in\mathcal{P}} f(\bar{x}_T, \mathbf{Q}) - \min_{x\in\mathcal{X}} f(x, \bar{\mathbf{Q}}_T) = O(T^{-1/2})$$

Dual DRO: tangent compression

Linearize loss pieces at the averaged decision:

$h_k(x, z) := \ell_k(\bar{x}_T, z) + \nabla_x \ell_k(\bar{x}_T, z)^\top (x - \bar{x}_T)$



$|\mathrm{supp}(\bar{\mathbf{Q}}_T)| \le T(N+1)$ Historical atoms ➡ **Tangent compression** LP for polyhedral $\mathcal{X}$ ➡ $|\mathrm{supp}(\widetilde{\mathbf{Q}})| \le N + n + 1$ Approximate dual solution

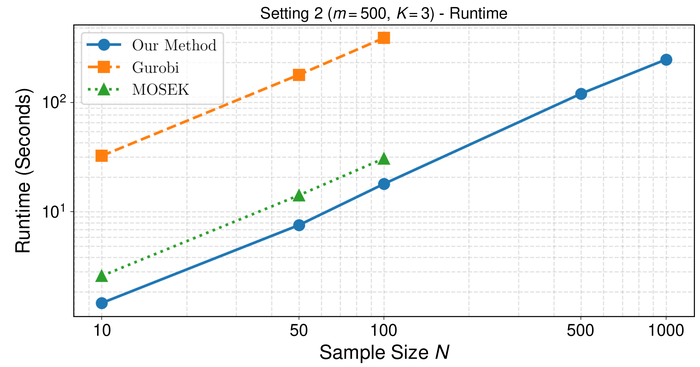
$$\mathrm{Gap}(\bar{x}_T, \widetilde{\mathbf{Q}}) \le \Delta_T + D\sqrt{2\kappa\Delta_T} \qquad \kappa = 0: \quad \mathrm{Gap}(\bar{x}_T, \widetilde{\mathbf{Q}}) \le \Delta_T$$

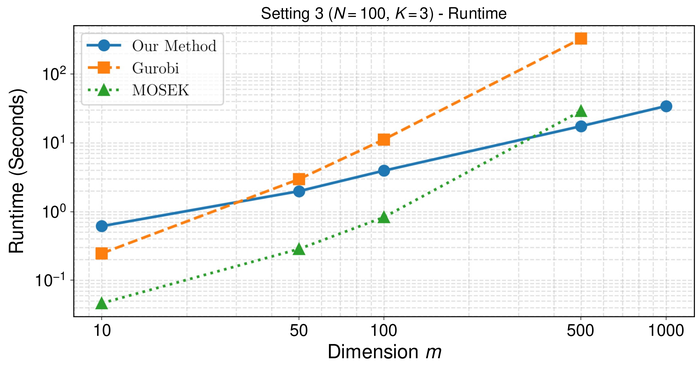
$\Delta_T := \mathrm{Gap}(\bar{x}_T, \bar{\mathbf{Q}}_T)$ D: diameter of $\mathcal{X}$; κ: smoothness of ℓ_k

Numerical experiments

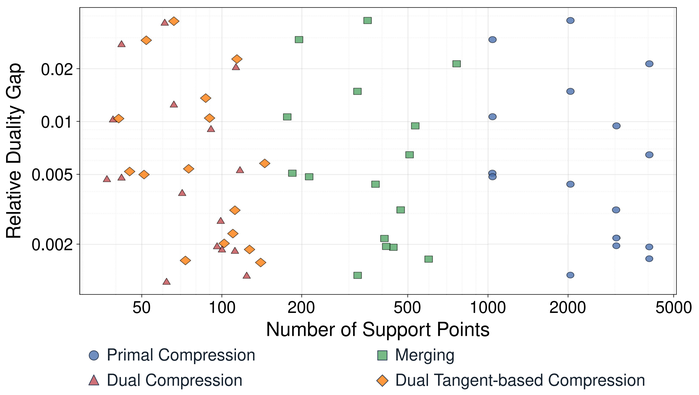
Piecewise quadratic losses with Euclidean transport cost ($\rho = 0.1$).

$$\ell_k(x, z) = x^\top C_k x + z^\top B_k x - z^\top A_k z, \qquad A_k, C_k \succ 0.$$





Relative error $\approx 10^{-5}$ vs. MOSEK; missing: out of memory or 600 seconds.



16 instances: $N, n \in \{25, 50, 75, 100\}$; $K = 3$, $T = 50$.

Selected references

P. Mohajerin Esfahani and D. Kuhn (2018). Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations. Mathematical Programming 171(1), 115-166.

R. Gao and A. Kleywegt (2023). Distributionally robust stochastic optimization with Wasserstein distance. Mathematics of Operations Research 48(2), 603-655.

D. Kuhn, S. Shafiee and W. Wiesemann (2025). Distributionally robust optimization. Acta Numerica 34, 579-804.

Guixian Chen gxchen@umich.edu