Two-Stage Distributionally Robust Optimization in Fixed-Charge Transportation

Mohamed El Tonbari, Alejandro Toriello, George Nemhauser

Georgia Institute of Technology

Introduction

In stochastic optimization, it is common to minimize the expected cost, but the probability distribution is often unknown in many applications, leading to techniques where the expected cost is approximated. One such method is the **Sample Average Approximation (SAA)**, where the expected cost is approximated by a set of observed samples. We are motivated by applications where historical data is scarce, in which case SAA can be too optimistic and lead to poor out-of-sample performance. More specifically, we are interested in problems arising in natural disaster management, where demand and cost data are limited, making SAA ill-suited. We focus on **hurricane disasters** in this work.

We turn to a **two-stage distributionally robust optimization** (TSDRO) model using a **Wasserstein** ambiguity set. The presence of binary variables in both stages of our model breaks convexity, and methods developed for the convex case with continuous support can only be used as an approximation (see [1, 2, 4]). We develop a column and constraint generation algorithm where we leverage the structure of the second stage value function and support set to efficiently solve the TSDRO.

Preliminaries

We are given a set of facilities I with capacities C_i, a set of demand nodes J, and a finite set of demand scenarios $\Xi = \{d^1, d^2, \ldots, d^R\}$.

First stage notation:

- x_i: 1 if facility i is built at a cost of O_i
- s_i: amount of resources to allocate to facility i at a cost of c

Second stage notation:

- y_{ij}: 1 if arc (i, j) is used for a fixed-charge cost f_{ij}
- t_{ij}: amount transported from i to j at cost v_{ij}
- u_j: unsatisfied demand at node j penalized by U

Two-Stage Model

The two-stage stochastic program can be written as

$$\begin{aligned} \min_{x,\, s} \quad & \sum_i [O_i x_i + c s_i] + \mathbb{E}_{\mathbb{P}}[Q(s, d^r)] \\ \text{s.t.} \quad & s_i \le C_i x_i, \quad \forall i \in I, \\ & s \ge 0, \\ & x \in \{0,1\}^{|I|} \end{aligned}$$

where

$$\begin{aligned} Q(s, d) = \min_{y,\, t,\, u} \quad & \sum_{i \in I, j \in J} [f_{ij} y_{ij} + v_{ij} t_{ij}] + U \sum_{j \in J} u_j \\ \text{s.t.} \quad & \sum_{j \in J} t_{ij} \le s_i, \qquad \forall i \in I, \\ & \sum_{i \in I} t_{ij} + u_j \ge d^r_j, \quad \forall j \in J, \\ & t_{ij} \le M y_{ij}, \qquad \forall i \in I, \quad \forall j \in J, \\ & y \in \{0,1\}^{|I| \times |J|}, \\ & t, u \ge 0. \end{aligned}$$

This model can be modified as necessary, such as adding a budget on the first stage costs or on the number of facilities that can be built.

Our **support set** is defined in a hierarchical fashion using four components:

1. $\xi_\ell \in L$: coordinates of the disaster's **landfall**
2. $\xi_c \in C$: **radius** of affected nodes
3. $\xi_p \in P$: **path** of the disaster
4. $\xi_f \in F$: **fraction of the population** affected at the landfall node, determining the intensity of the disaster.

A scenario $\xi^r = (\xi^r_\ell, \xi^r_c, \xi^r_p, \xi^r_f)$ determines the demands d^r_j of each node $j \in J$.

Assumptions:

(A1) There can only be one landfall.

(A2) Given ξ_f, the intensities (i.e. the fractions of the population affected) at the remaining affected nodes within radius ξ_c are deterministic, decreasing the further they are from the landfall.

Distributionally Robust Optimization

In DRO, we seek to minimize the **worst case expected cost** with respect to probability distributions belonging to the **Wasserstein ambiguity set**.

$$\mathbb{E}_{\mathbb{P}}[Q(s, \xi^r)] \Rightarrow \max_{\mathbb{P} \in \mathcal{B}_W(\mathbb{P}_N, \theta)} \mathbb{E}_{\mathbb{P}}[Q(x, \xi)]$$

where $\mathcal{B}_W(\mathbb{P}_N, \theta) = \{\mathbb{P} : \mathrm{d}_W(\mathbb{P}, \mathbb{P}_N) \le \theta\}$ is the Wasserstein ball of radius θ centered at the empirical distribution $\mathbb{P}_N$ constructed from N samples. DRO can be seen as a **generalization of SAA and robust optimization (RO)**.

$$\text{SAA} \; \underset{0 \qquad\qquad \theta \qquad\qquad \infty}{\longrightarrow} \; \text{RO}$$

The **Wasserstein distance** $\mathrm{d}_W(\cdot, \cdot)$ between two probability measures is the optimal transport of weights between scenarios such that the two measures "match", at a cost of the distance between the scenarios. In Figure 1, the colors represent the **optimal transport from distribution p to q**, and the only cost incurred is moving a **weight of 0.1 from ξ_1 to ξ_2 at a cost of $\mathrm{d}(\xi_1, \xi_2)$**, where $\mathrm{d}(\cdot, \cdot)$ is any valid metric.

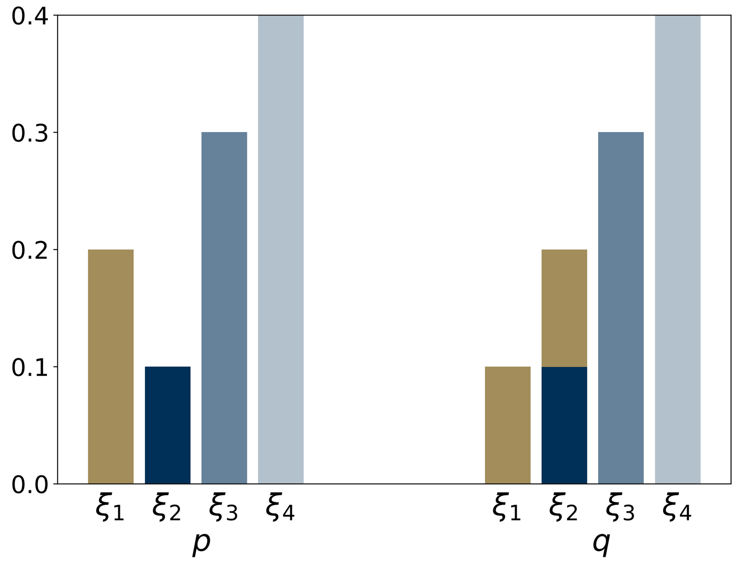

Fig. 1: Wasserstein Distance Example

Column & Constraint Generation Algorithm

We consolidate all first and second stage decisions into x and y. Given scenario ξ^r and first stage decision x, let $Y(x, \xi^r)$ be the feasible region of the second stage fixed-charge transportation problem. Given a set of N samples $\{\hat{\xi}^1, \ldots, \hat{\xi}^N\}$, an **extensive reformulation** of our TSDRO model is

$$\begin{aligned} \min_{\substack{x, y \\ \lambda \ge 0}} \quad & c^\top x + \theta\lambda + \frac{1}{N}\sum_{n=1}^{N} \alpha_n \\ \text{s.t.} \quad & \alpha_n \ge q^\top y^r - \lambda\, \mathrm{d}(\xi^r, \hat{\xi}^n), \quad r = 1, \ldots R, \quad n = 1, \ldots N, \qquad (1a) \\ & y^r \in Y(x, \xi^r), \qquad r = 1, \ldots, R, \qquad (1b) \\ & x \in X \end{aligned}$$

- Number of scenarios R is very large, and solving (1) is infeasible. We instead solve (1) with a small subset of scenarios $\hat{R} \subseteq R$.
- Given optimal solutions to the restricted master $\hat{x}$ and $\hat{\lambda}$, we add a new scenario $r \in R \setminus \hat{R}$ as follows:
  - (i) Solve $Q(\hat{x}, \xi^r)$. Let y^* be the optimal solution.
  - (ii) If $\alpha_n < q^\top y^* - \hat{\lambda}\, \mathrm{d}(\xi^r, \hat{\xi}^n)$ for any sample n, add new set of constraints (1a) and (1b) associated with r and n, and new variables y^r.
- We leverage the structure of $Q(x, \xi)$ and our support set Ξ, identifying dominated scenarios which need not be considered, and enumerating scenarios more efficiently.

Value Function Structure

Distance Metric $\mathrm{d}(\cdot, \cdot)$:

- Can be the Euclidean distance between demand vectors, for example. Two demand vectors can, however, be close to each other, but result from very different hurricanes in different locations in the network.
- Instead, we define $\mathrm{d}(\cdot, \cdot)$ as the **Euclidean distance between vectors $(\xi_\ell, \xi_c, \xi_p, \xi_f)$**, i.e. the random vectors which define our support set.

This permits us to better capture the distance between two disaster scenarios and leads to the following two results.

Theorem 1 (Dominated Scenarios). *For each sample $\hat{\xi}^n$, only considering scenarios $\xi \in \Xi_n \subseteq \Xi$, where*

$$\begin{aligned} \Xi_n = \{\xi \in \Xi : & \xi_f > \hat{\xi}^n_f, \ \forall \xi \text{ where } \xi_c < \xi^n_c, \\ & \xi_c > \hat{\xi}^n_c, \ \forall \xi \text{ where } \xi_f < \xi^n_f\}. \end{aligned}$$

leads to an equivalent problem.

(A3) Facilities can only serve a limited number k_i of demand nodes, and are not limited by an amount pre-allocated in the first stage.

Theorem 2 (Concavity of Value Function). *Assuming **(A3)** and given $\hat{x}$, the value function $Q(\hat{x}, \xi)$ is piece-wise concave with respect to the intensity of the disaster ξ_f. Moreover, $Q(\hat{x}, \xi) - \hat{\lambda}\, \mathrm{d}(\xi, \hat{\xi}^n)$ is concave with respect to ξ_f for any sample $\hat{\xi}^n$ and fixed $\hat{\lambda}$.*

- Goal is to maximize $g(\xi) = Q(\hat{x}, \xi) - \hat{\lambda}||\xi - \hat{\xi}^n||$ with respect to ξ for each sample given $(\hat{x}, \hat{\lambda})$.
- Instead of enumerating $\mathcal{O}(|\boldsymbol{L}| \times |\boldsymbol{C}| \times |\boldsymbol{P}| \times |\boldsymbol{F}|)$ scenarios, we can perform a **Fibonacci** search on F for each landfall, radius and possible path, enumerating $\mathcal{O}(|\boldsymbol{L}| \times |\boldsymbol{C}| \times |\boldsymbol{P}| \times \log|\boldsymbol{F}|)$ instead.

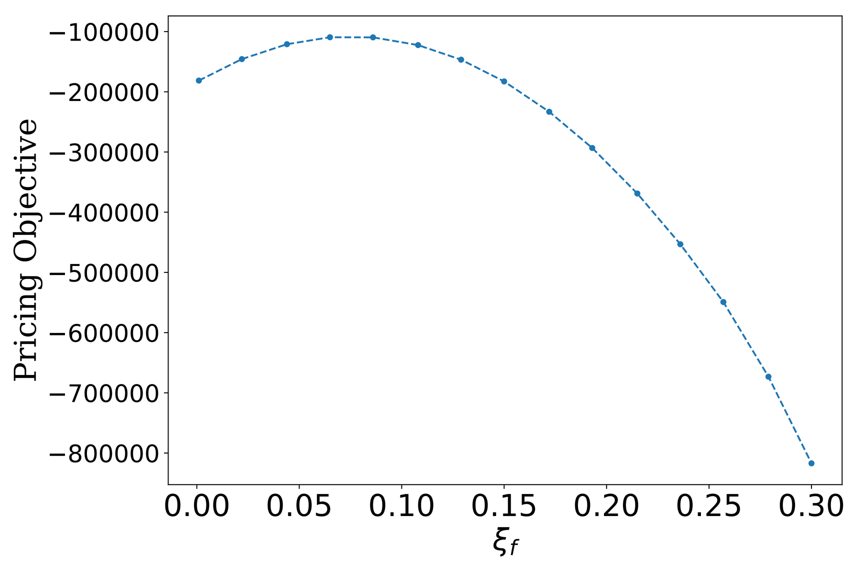

Fig. 2: Pricing objective $g(\xi)$ as a function of ξ_f

Without **(A3)**, Theorem 2 does not hold in general, but we observe concavity in most cases in our experiments, and using Fibonacci search still leads to speedups.

Fibonacci Search Example Let $f(\xi_f) = g(\cdot, \cdot, \cdot, \xi_f)$. In Figure 3, we perform one step of Fibonacci search, and maintain upper and lower bounds UB and LB. In our example, $f(\xi_{f_1}) \ge f(\xi_{f_2})$, and we update ξ_{f_1}, ξ_{f_2}, and UB accordingly.

Note: indices of LB, ξ_{f_1}, ξ_{f_2} and UB are always **Fibonacci numbers**.

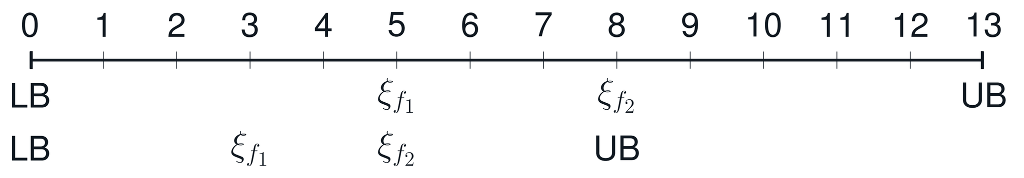

Fig. 3: One Step of Fibonacci Search

Experiments

We present preliminary computational results. In Figure 4, we compare the runtimes of three different enumeration techniques:

1. enumerate all scenarios and add the scenario which violates (1a) the most
2. enumerate and stop at the first scenario which violates (1a)
3. enumerate landfalls, radii, and paths, and perform Fibonacci search on F

In Figure 5, we plot the opened facilities resulting from SAA and DRO using a case study on hurricane threats in the Gulf of Mexico states, using data from [3].

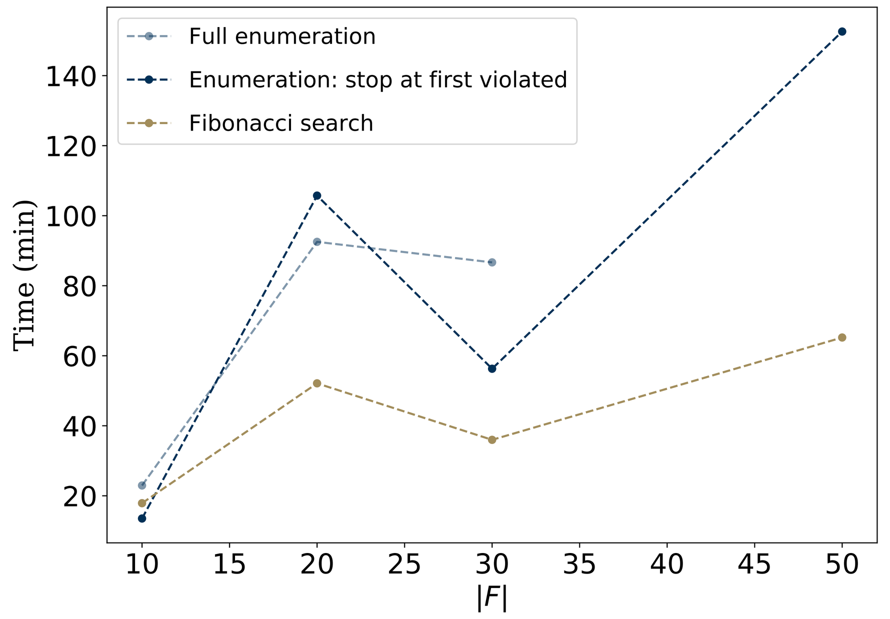

Fig. 4: Runtime comparison of different enumeration techniques

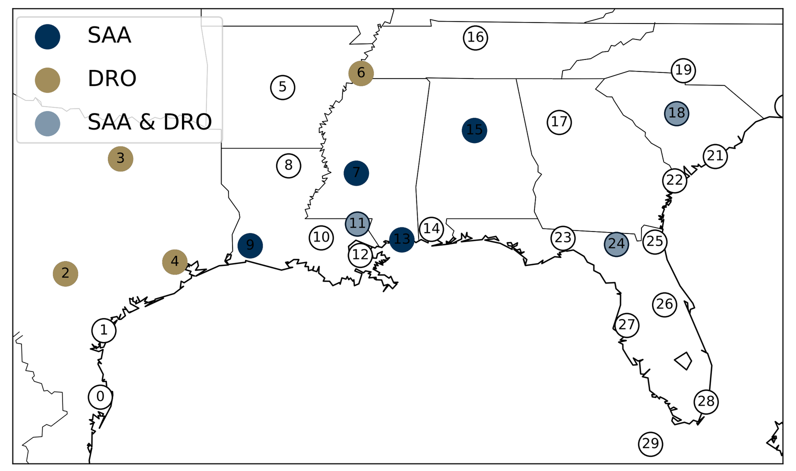

Fig. 5: Open facilities resulting from SAA and DRO: a case study on the Gulf of Mexico states

References

[1] Peyman Mohajerin Esfahani and Daniel Kuhn. "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations". In: *Mathematical Programming* 171.1-2 (2018), pp. 115–166.

[2] Rui Gao and Anton J Kleywegt. "Distributionally robust stochastic optimization with Wasserstein distance". In: *arXiv preprint arXiv:1604.02199* (2016).

[3] Carmen G Rawls and Mark A Turnquist. "Pre-positioning of emergency supplies for disaster response". In: *Transportation research part B: Methodological* 44.4 (2010), pp. 521–534.

[4] Weijun Xie. "Tractable Reformulations of Distributionally Robust Two-stage Stochastic Programs with ∞-Wasserstein Distance". In: *arXiv preprint arXiv:1908.08454* (2019).

Acknowledgement

This resesarch has been supported by the Office of Naval Research under the Award number N00014-18-1-2075.