A Literature Review on Data-Driven Distributed Robust Optimization using Wasserstein Metric

Yuyang Ma *yuyang.ma@lehigh.edu*
Department of Industrial and System Engineering
Lehigh University

Introduction

Distributionally robust optimization (DRO) has gained popularity in recent years due to its ability to address uncertainty in optimization problems. Compared to DRO, stochastic programming assumes that the distribution of the random parameters is completely known. However, in most real-world cases, only a series of historical data is available, which cannot stands for the true distribution of random parameters we want to estimate. It means the solution obtained from stochastic programming methods can be biased and sub-optimal for the true problem. In this case, distributionally robust optimization, which has the capability to handle the ambiguity of distribution, has attracted attention recently. The three papers in this review all contribute to the evolving field of DRO, particularly focusing on the use of the Wasserstein metric to measure distributional uncertainty. This literature review will provide an overview of these papers, highlighting their key concepts and contributions while also discussing their interconnections.

Mathematical Background

Making decision under uncertainty is always challenging, and decison-makers usually face a two-stage decision problem. Stochastic programming (SP) is an effective method to address uncertainty. The traditional two-stage stochastic optimization can be described as

$$\text{(SP)} \quad \min_{x \in X} \; c^T x + \mathbf{E}_{\mathbb{P}} \left[\mathcal{Q}(x, \xi) \right], \tag{1}$$

where x is the variable of first-stage decisions, and $X \subseteq \mathbb{R}^n$ is a convex set representing the feasible region for the first-stage decision variable. The term $\mathbf{E}_{\mathbb{P}} \left[\mathcal{Q}(x, \xi) \right]$ stands for the expectation value of the second stage cost $\mathcal{Q}$ when assuming the probability distribution of Ξ to be $\mathbb{P}$. Let ξ be a realization of random variable $\Xi \in \mathbb{R}^m$, which is the set stands for the potential realities after making first-stage decision x. For a certain x and ξ, the corresponding second stage cost is

$$\mathcal{Q}(x, \xi) = \min_{y \in Y} \left\{ d^T y : \; A(\xi)x + B(\xi)y \geq b(\xi) \right\}.$$

Similar to the first-stage decision variable, y here is the second-stage decision variable, and $Y \in \mathbb{R}^m$ stands for the feasible region of y.

In SP's framework 1, ξ is defined on a probability space $(\Omega, \sigma(\Omega), \mathbb{P})$, where Ω is the sample space of ξ, $\sigma(\Omega)$ is the σ-algebra of Ω, and $\mathbb{P}$ is the probability measure Zhao & Guan (2018). Under this setting, probability distribution $\mathbb{P}$ is given, and to solve the problem, the sample average approximation method is usually applied in which the random parameters are captured by a number of scenarios sampled from the true distribution. This approach becomes computationally heavy when the number of scenarios increases.

Considering the disadvantages of stochastic optimization approaches, plus the fact that the true distribution of the random parameters is usually unknown, it is hard to predict accurately. Moreover, any inaccurate estimation of the true distribution may lead to biased solutions and make the solutions sub-optimal. In order to solve this difficulty, distributionally robust optimization approaches are proposed and can be described as follows:

$$(\mathrm{DR-SP/DRO}) \quad \min_{x \in X} \; c^T x + \max_{\hat{\mathbb{P}} \in \mathcal{D}} \mathbf{E}_{\hat{\mathbb{P}}} \left[\mathcal{Q}(x, \xi) \right]. \tag{2}$$

Instead of having the probability distribution $\mathbb{P}$ by assumption, DRO 2 allows the ambiguity of $\mathbb{P}$. The $\mathcal{D}$ here is a confident set ensuring that the true distribution $\mathbb{P}$ is within this set based on a statistical inference. The purpose of DRO is to minimize the worst-case problem of any possible $\hat{\mathbb{P}}$ in confident set $\mathcal{D}$. Based on the historical data, an empirical distribution can be used as the reference distribution. Then $\mathcal{D}$ can be constructed by using any metrics to calculate the distance between true probability distribution and reference distribution. Let θ be the distance between true distribution $\mathbb{P}_0$ and estimated distribution $\hat{\mathbb{P}}$ under metric M, then the confident set $\mathcal{D}$ can be described as following:

$$\mathcal{D} = \left\{ \hat{\mathbb{P}} : \; d_M \left(\mathbb{P}_0, \hat{\mathbb{P}} \right) \leq \theta \right\}$$

The Wasserstein distance is a distance function defined between probability distributions on a given metric space W. iven two probability distributions $\mathbb{P}$ and $\hat{\mathbb{P}}$ on the supporting space Ω, the Wasserstein metric is defined as

$$d_W \left(\mathbb{P}, \hat{\mathbb{P}} \right) := \inf_{\pi} \left\{ \mathbf{E}_\pi[\rho(X, Y)] : \mathbb{P} = \mathcal{L}(X), \hat{\mathbb{P}} = \mathcal{L}(Y) \right\}$$

where $\rho(X, Y)$ is defined as the distance between random variables X and Y, and the infimum is taken over all joint distributions π with marginals $\mathbb{P}$ and $\hat{\mathbb{P}}$.

Literature Review

The first paper Delage & Ye (2010) explores the intersection of distributionally robust optimization and moment uncertainty. It highlights the application of DRO in data-driven problems, emphasizing the modeling of uncertainty through moments (e.g., mean and variance). This approach provides a different perspective on distributional uncertainty and offers insights into its implications for decision-making in various domains.

The second paper Esfahani & Kuhn (2015) introduces the Wasserstein metric as a powerful tool for formulating distributionally robust optimization problems. It emphasizes performance guarantees and tractable reformulations, offering insights into the trade-offs between conservatism and optimality when dealing with uncertainty in data. This paper demonstrates that the worst-case expectation over a Wasserstein ambiguity set can be computed efficiently via convex optimization techniques for numerous loss functions of practical interest. Furthermore, authors propose an efficient procedure for constructing an extremal distribution that attains the worst-case expectation—provided that such a distribution exists. The paper's focus on computational efficiency and robustness makes it a significant contribution to the DRO literature.

In Zhao & Guan (2018), they extend the concept of data-driven optimization using the Wasserstein metric by incorporating risk aversion. It addresses the practical challenge of optimizing decisions under uncertainty while considering risk preferences. The integration of risk measures with the Wasserstein metric enhances the applicability of DRO to scenarios where decision-makers seek to balance robustness and risk management.

These three papers share a common thread in their utilization of the Wasserstein metric as a means to model distributional uncertainty. The first paper introduces this concept and focuses on performance guarantees and tractable reformulations, the second extends it to incorporate risk aversion, addressing a broader set of decision-making scenarios. The third paper, on the other hand, explores moment uncertainty within the DRO framework, shedding light on the flexibility of DRO models in handling various forms of uncertainty. These three papers collectively contribute to the advancement of data-driven distributionally robust optimization using the Wasserstein metric. They offer valuable insights into modeling and addressing uncertainty in optimization problems, and their interconnections highlight how powerful the DRO framework is when handling uncertainty in real world problems. As the field of DRO continues to evolve, these papers provide a strong foundation for future research in data-driven robust decision-making under uncertainty.

References

Erick Delage and Yinyu Ye. Distributionally robust optimization under moment uncertainty with application to data-driven problems. *Operations Research*, 58:595–612, 2010.

Peyman Mohajerin Esfahani and Daniel Kuhn. Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations. *Mathematical Programming*, 171:115 – 166, 2015. URL `https://api.semanticscholar.org/CorpusID:14542431`.

Chaoyue Zhao and Yongpei Guan. Data-driven risk-averse stochastic optimization with wasserstein metric. *Operations Research Letters*, 46:262–267, 2018.