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

Abstract

Optimization under uncertainty is undergoing a transformative evolution with the advent of distributionally robust optimization (DRO), a programming that addresses the challenges by ambiguous data distributions. Rooted in the acknowledgment of imperfect models and limited data, DRO seeks robust solutions by optimizing over a set of potential distributions rather than assuming a certain distribution. This literature review delves into the intersection of DRO and the Wasserstein metric, a measure capturing distributional dissimilarity. The introduction focuses the significance of DRO and the Wasserstein metric in handling uncertainty, setting the stage for a comprehensive exploration. The mathematical background illuminates the key concepts underpinning DRO and the Wasserstein metric, establishing a foundation for the subsequent review. The literature review encompasses three papers, each leveraging the Wasserstein metric within the DRO framework. The work Esfahani & Kuhn (2015) introduces the metric, emphasizing performance guarantees and tractable reformulations. The second Zhao & Guan (2018) extends the approach to incorporate risk aversion, broadening the scope of DRO to scenarios involving risk preferences. The third paper Delage & Ye (2010) explores DRO under moment uncertainty, shedding light on the flexible application of the framework to data-driven problems. All these contributions advance our understanding of data-driven distributionally robust optimization, highlighting the Wasserstein metric's role in enhancing robust decision-making under diverse forms of uncertainty.

Introduction

Optimization under uncertainty stands as a paramount challenge in decision-making across diverse fields, from finance and operations research to engineering and environmental science. The conventional optimization frameworks, often relying on deterministic or probabilistic assumptions, frequently fall short in capturing the complexities inherent in real-world data and models. This realization has encouraged the evolution of distributionally robust optimization (DRO), a paradigm that offers a robust and adaptive approach to decision-making when facing uncertainty.

In the realm of finance, where market conditions are uncertain and subject to sudden changes, DRO has found practical applications. Researchers have leveraged DRO methodologies to formulate robust portfolio optimization strategies, ensuring that investment decisions remain resilient across various market scenarios Ben-Tal & Nemirovski (2013). The DRO framework proves invaluable in scenarios where traditional models may fail to capture the true distribution of asset returns, providing a more realistic and robust foundation for financial decision-makers.

In the context of supply chain management, where disruptions are commonplace, DRO has been applied to develop resilient logistics strategies. The work of Bertsimas & Sim (2004) exemplifies the use of robust optimization in supply chain planning, considering uncertainty in demand and supply chain parameters. Such applications emphasizes the practical relevance of DRO in optimizing decision-making processes for complex systems.

The first crucial step in understanding DRO lies in acknowledging the uncertainty inherent in the data. Whether it be due to limited sample sizes, model misspecification, or the variability of real-world systems, uncertainty introduces a layer of complexity that traditional optimization approaches struggle to address. DRO, in contrast, takes a more solid approach, seeking solutions that perform well across a range of potential distributions, providing a robust foundation for decision-making. The Wasserstein metric within the DRO framework rooted in optimal transport theory, captures the underlying geometry of probability distributions, offering a more rigorous measure of dissimilarity than traditional metrics. As Wasserstein-based DRO gains prominence, its real-world applications span diverse domains, from finance and logistics to energy and healthcare.

Building on these foundations, Esfahani & Kuhn (2015) contributes significantly to this evolving landscape. By introducing the Wasserstein metric into the DRO framework, the paper not only enhances the robustness of optimization solutions but also provides guarantees on their performance. Moreover, the emphasis on tractable reformulations is a crucial step toward making DRO computationally feasible for real-world problems. As we navigate through the subsequent sections, we will firstly delve into the mathematical background of both DRO and Wasserstein metrics. Then we discuss key insights and findings offered by each paper.

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

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

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:

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

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. Moment uncertainty is a critical aspect explored in the broader DRO framework, and it refers to uncertainty regarding statistical moments beyond the first order. The moments of a distribution, such as mean, variance, and skewness, capture essential features that define the shape and characteristics of the distribution. In the context of DRO, the first paper highlights the application of DRO in data-driven problems, emphasizing the modeling of uncertainty through moments. This paper 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. In the context of optimization under uncertainty, risk aversion refers to the preference for strategies that minimize potential losses, even at the expense of reduced gains. Decision-makers with a risk-averse stance are inclined to prioritize robustness and stability, demonstrating a lower tolerance for exposure to uncertain and potentially adverse outcomes. Within the DRO framework, the integration of risk aversion involves incorporating risk measures alongside the Wasserstein metric. This extension allows decision-makers to strike a balance between robustness and risk tolerance, catering to scenarios where mitigating downside risks is of extreme significance. The second paper thus contributes to a more comprehensive understanding of how Wasserstein-based DRO models can align with the nuanced risk preferences of decision-makers in various industries. It addresses the practical challenge of optimizing decisions under uncertainty while considering risk preferences and it allows 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

Aharon Ben-Tal and Arkadi Nemirovski. Robust optimization—methodology and applications. *Mathematical Programming*, 92(3):453–480, 2013.

Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004.

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.