Tightening ambiguity set characterizations for data-driven distributionally robust optimization

Lotfi Mustapha Chaouach[1], Dimitris Boskos[1], and Tom Oomen[1,2]

[1]Delft Center for Systems and Control, 2628CD, Delft, the Netherlands (e-mail: l.chaouach@tudelft.nl)
[2]Eindhoven University of Technology, the Netherlands

1 Introduction

Real-world systems are always uncertain. Their increasing complexity, sophistication, and connectivity generate further sources of uncertainty that is characterized by deterministic (worst-case) or stochastic models. The latter assume existence of a probability distribution, which is typically inferred from data. To build such data-driven models, we need to account for limitations induced by the fact that data may be corrupted or only available at small amounts. These aspects influence the accuracy of inferences about the underlying probability distributions and their usefulness for decision making. To overcome these issues, it is required to derive reliable probabilistic models of the uncertain components from the available data under the least possible amount of conservativeness.

2 Problem formulation

The central problem in stochastic optimization is to take optimal decisions in problems affected by randomness. A typical stochastic optimization problem has the form

$$\inf_{x\in\mathscr{X}} \mathbb{E}_{P_\xi}\left[f(x,\xi)\right] \quad (1)$$

where f is the objective function, x is the decision variable, and $\xi \in \mathbb{R}^d$ is a random variable with distribution P_ξ. In practice, as P_ξ is often unknown, it is approximated by the empirical distribution $P_\xi^N := \sum_{i=1}^N \delta_{\hat{\xi}^i}$ of i.i.d. samples. Yet, for small amounts of data, this approach may become insufficient, as the approximation P_ξ^N may exhibit significant deviations from P_ξ. To address this issue, uncertainty in the distribution is considered as

$$\inf_{x\in\mathscr{X}} \max_{P_\xi\in\mathscr{P}^N} \mathbb{E}_{P_\xi}\left[f(x,\xi)\right], \quad (2)$$

which is a distributionally robust optimization problem [1, 2, 3], where $\mathscr{P}^N$ is an ambiguity set of distributions that is informed by the samples and contains plausible models for the true distribution. A convenient choice of ambiguity sets for such problems are balls in the Wasserstein metric centered at the empirical distribution P_ξ^N. For compactly supported distributions, choosing the radius

$$\varepsilon_N(\beta,\rho) := \left(\frac{\ln(C\beta^{-1})}{c}\right)^{\frac{1}{q}} \frac{\rho}{N^{\frac{1}{q}}}, \quad q := \max\{2p,d\}, \quad (3)$$

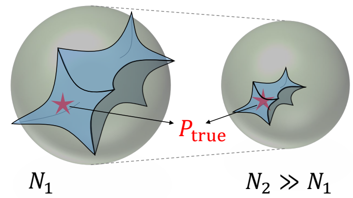

Figure 1: High-dimensional hyperrectangles shrink much faster with the number of samples compared to Wasserstein ambiguity balls.

where $d \neq 2p$ and ρ is the diameter of the support of P_ξ, guarantees that the ambiguity ball contains the true distribution with probability at least $1-\beta$. However, (3) implies that for high-dimensional random variables, the radius decreases with the excessively slow rate of the order of $N^{-\frac{1}{d}}$.

3 Ambiguity hyperrectangles

To address the conservative decrease rate in (3), we exploit independence of lower-dimensional components of the random variable $\xi = (\xi_1,\dots,\xi_n)$. We build a lower-dimensional ambiguity ball for each component and construct an ambiguity hyperrectangle by taking all product measures across the individual distributions from the n balls. Such a rectangle shrinks at the faster rate $N^{\frac{-1}{q^\star}}$ with the number of samples, where $q^\star := \max\{q_k, k=1,\dots,n\}$, $q_k := \max\{2p,d_k\}$ for $d_k \neq 2p$ and d_k is the dimension of ξ_k. The ambiguity hyperrectangle is much smaller than the original ambiguity ball and contains the true distribution with the same confidence (cf. Figure 1).

References

[1] P. M. Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations," Mathematical Programming, vol. 171, no.1-2, pp. 115–166, 2018.

[2] R. Gao and A. Kleywegt, "Distributionally robust stochastic optimization with Wasserstein distance,"arXiv preprint arXiv:1604.02199, 2016.

[3] J. Blanchet and K. Murthy, "Quantifying distributional model risk via optimal transport," Mathematics of Operations Research, vol. 44, no. 2, pp. 565–600, 2019.