Proceedings of the 2019 Winter Simulation Conference
N. Mustafee, K.-H.G. Bae, S. Lazarova-Molnar, M. Rabe, C. Szabo, P. Haas, and Y.-J. Son, eds.

VALIDATING OPTIMIZATION WITH UNCERTAIN CONSTRAINTS

Henry Lam
Huajie Qian

Department of Industrial Engineering and Operations Research
Columbia University
500 W. 120th Street
New York, NY 10027, USA

ABSTRACT

We consider optimization with uncertain or probabilistic constraints under the availability of limited data or Monte Carlo samples. In this situation, the obtained solutions are subject to statistical noises that affect both the feasibility and the objective performance. To guarantee feasibility, common approaches in data-driven optimization impose constraint reformulations that are "safe" enough to ensure solution feasibility with high confidence. Often times, selecting this safety margin relies on loose statistical estimates, in turn leading to overly conservative and suboptimal solutions. We propose a validation-based framework to balance the feasibility-optimality tradeoff more efficiently, by leveraging the typical low-dimensional structure of solution paths in these data-driven reformulations instead of estimates based on the whole decision space utilized by past approaches. We demonstrate how our approach can lead to a feasible solution with less conservative safety adjustment and confidence guarantees.

1 INTRODUCTION

We consider stochastically constrained optimization problems in the form

$$\begin{array}{ll} \min\limits_{x\in\mathscr{X}} & f(x) \\ \text{subject to} & H(x) := \mathbb{E}_F[h(x,\xi)] \geq \gamma \end{array} \tag{1}$$

where $x \in \mathbb{R}^d$ is the decision variable, $\mathscr{X} \subset \mathbb{R}^d$ is the deterministic decision space, $\xi \in \mathbb{R}^m$ is a random vector following an unknown distribution F, and $\mathbb{E}_F[\cdot]$ and correspondingly $P_F(\cdot)$ denotes the expectation and probability under F. $h(\cdot,\cdot) : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$ is a known function, which represents some notion of gain whose expected value is constrained above a given threshold γ. We assume $f(\cdot)$ to be a deterministic function, as our focus is on handling the stochasticity in the constraint (this assumption can be relaxed). Formulation (1) appears as chance constrained programs (CCPs) (Prékopa 2003) when $h(x,\xi) := \mathbf{1}((x,\xi) \in A)$, in which case we want the solution to satisfy the event with high probability. It also appears in convex expected value constrained programs (Atlason et al. 2004; Krokhmal et al. 2002) when $h(x,\xi)$ is concave in x for every ξ.

We focus on the situation where the governing distribution F is not fully known but only observed through a set of i.i.d. data or Monte Carlo sample $\{\xi_1,\ldots,\xi_n\}$. The goal is to compute a solution, based on the available data, that is both feasible and possesses good objective performance. Note that, due to the statistical noise from the data, feasibility can be guaranteed at best with a high confidence, which in turn also imposes a price on the optimality of the obtained solution. Finding a good solution thus requires striking a balance between feasibility and optimality: If a procedure is overly protective (i.e., removing much of the feasible region), then the obtained solution will be surely feasible, but optimality will suffer,

and vice versa. The efficacy of this balancing depends on the efficiency in assimilating the data into the optimization and in the corresponding estimation procedure. The aim of this paper is to construct a theoretically sound and implementable framework that makes such balancing highly efficient.

2 EXISTING APPROACHES AND CHALLENGES

Before introducing our framework, let us first discuss the established approaches and point out the major challenges for this problem, which has been studied prominently in the data-driven optimization literature. The common practice is to reformulate the unknown constraint in (1) into a data-driven constraint that depends only on the data, so that this reformulated constraint is "safe". To be more precise, let $\mathscr{F}$ be the feasible region of (1). We would like to construct a region $\hat{\mathscr{F}}$ that depends only on $\{\xi_1,\ldots,\xi_n\}$, such that an optimal solution $\hat{x}^*$ (or a best obtainable feasible solution to (2) if exact computation is hard) of the problem

$$\begin{array}{ll} \min\limits_{x\in\mathscr{X}} & f(x) \\ \text{subject to} & x\in\hat{\mathscr{F}} \end{array} \tag{2}$$

satisfies

$$P_{data}(\hat{x}^*\in\mathscr{F})\geq 1-\beta \tag{3}$$

where P_{data} refers to the probability generated from the data. In other words, the data-driven solution $\hat{x}^*$ is indeed feasible for the original problem, with a confidence level $1-\beta$.

Note that simply replacing $H(\cdot)$ with a naive point estimate (e.g., the sample mean from the data) is typically inadequate to guarantee (3). To see this, suppose the true optimal solution x^* is at the boundary of $\mathscr{F}$, i.e., $H(x^*)=\gamma$. If we consider $\hat{\mathscr{F}}=\{x:(1/n)\sum_{i=1}^n h(x,\xi_i)\geq\gamma\}$. Then, roughly speaking, with half probability the obtained solution $\hat{x}^*$ will have $H(\hat{x}^*)$ below γ, which leads to infeasibility for the original problem. This issue may not arise if x^* or $\hat{x}^*$ is in the interior of the feasible region. However, a priori we do not know our decision. Thus, in general, to guarantee (3), one would impose a safety margin on an estimation of $H(\cdot)$, such that any solution obtained from (2) is also feasible for (1), with the required confidence. That is, we want

$$P_{data}(\hat{\mathscr{F}}\subset\mathscr{F})\geq 1-\beta. \tag{4}$$

We contend that most approaches in data-driven optimization rely on the above reasoning and are based on (4). In particular, (4) provides a convenient way to certify feasibility, by requiring that *all* solutions feasible for (2) are also feasible for (1) with high confidence. This set-level guarantee generally leads to a simultaneous estimation problem across all x in the decision space $\mathscr{X}$, for which a proper control of the statistical error can lead to a substantial shrinkage of the size of $\hat{\mathscr{F}}$ that exacerbates with problem dimension (either of the decision space or the probability space).

We provide several examples to illustrate the phenomenon above. Our discussion considers the (single) CCP, where $H(x)$ is in the form $P_F(G(x,\xi)\leq b)$ with $G(x,\xi):\mathbb{R}^d\times\mathbb{R}^m\to\mathbb{R}$.

Sample average approximation (SAA): The SAA approach sets

$$\hat{\mathscr{F}}=\left\{x\in\mathscr{X}:\frac{1}{n}\sum_{i=1}^n\mathbf{1}(G(x,\xi_i)+\varepsilon\leq b)\geq\gamma+\delta\right\}$$

where ε and δ are suitably tuned parameters. For example, when G is Lipschitz continuous in x, selecting $\delta=\Omega(\sqrt{(d/n)\log(1/\varepsilon)})$ can guarantee (4) (Luedtke and Ahmed 2008), and similar relations also hold in discrete decision space (Luedtke and Ahmed 2008) and expected value constraints (Wang and Ahmed 2008). These estimates come from concentration inequalities in which union bounds are needed and give rise to the dependence on the dimension d. Note that the resulting margin δ scales in the order of $\sqrt{d}$, and also that to get any reasonably small δ n must be of higher order than d.

Robust optimization (RO) and safe convex approximation (SCA): RO sets

$$\hat{\mathscr{F}} = \{x \in \mathscr{X} : G(x,\xi) \le b, \text{ for all } \xi \in \mathscr{U}\} \tag{5}$$

where $\mathscr{U}$ is known as the uncertainty set, and ξ in (5) is viewed as a deterministic unknown (Bertsimas et al. 2011; Ben-Tal et al. 2009). A common example of $\mathscr{U}$ is an ellipsoidal set $\{\xi : (\xi-\hat{\mu})^T\hat{\Sigma}^{-1}(\xi-\hat{\mu}) \le \rho\}$ where $\hat{\mu} \in \mathbb{R}^d$, $\hat{\Sigma} \in \mathbb{R}^{d\times d}$ a positive semidefinite matrix, and $\rho \in \mathbb{R}$. Here the center $\hat{\mu}$ and shape $\hat{\Sigma}$ typically correspond to the mean and covariance of the data, and ρ controls the set size. A duality argument shows that, in the case of linear chance constraint in the form $G(x,\xi) = x^T\xi$, (5) is equivalent to the quadratic constraint $\hat{\mu}^T x + \sqrt{\rho}\|\hat{\Sigma}^{1/2}x\|_2 \le b$. Using such type of convex constraints as inner approximations for intractable chance constraints is also known as SCA (e.g., Nemirovski 2003; Nemirovski and Shapiro 2006).

If the random variable ξ is known to be bounded, the above approach guarantees the obtained solution has a satisfaction probability of order $1-e^{-\rho/2}$ via Hoeffding's inequality, and ρ is chosen by matching this expression with the tolerance level γ. Although ρ calibrated this way may not explicitly depend on the problem dimension, the level of conservativeness measured by the shrinkage of $\mathscr{F}$ to $\hat{\mathscr{F}}$ is governed by concentration bounds whose tightness can be challenging to quantify. Another viewpoint that has been utilized recently in data-driven RO (Bertsimas et al. 2018; Tulabandhula and Rudin 2014; Goldfarb and Iyengar 2003; Hong et al. 2017) is to take $\mathscr{U}$ to be a set that contains γ-content of the distribution of ξ, i.e., $P_F(\xi \in \mathscr{U}) \ge \gamma$, with a confidence level $1-\beta$. In this case, a solution $\hat{x}^*$ obtained from solving (5) would satisfy $P_F(G(\hat{x}^*,\xi) \le b) \ge \gamma$ with at least $1-\beta$ confidence, thus achieving (4) as well. Such generated uncertainty set however typically has a size that scales with the dimension of the probability space. For example, consider $G(x,\xi) = x^T\xi$ with $\xi \in \mathbb{R}^m$ being standard multivariate Gaussian and the uncertainty set $\mathscr{U}$ is an ellipsoid with $\hat{\mu}$ and $\hat{\Sigma}$ being the true mean and covariance, i.e., $\mathscr{U} = \{\xi \in \mathbb{R}^d : \|\xi\|_2^2 \le \rho\}$. Then, in order to make $\mathscr{U}$ a γ-content set the radius ρ has to be at least of order m since $\|\xi\|_2^2$ has a mean m, resulting in the robust counterpart $\sqrt{\rho}\,\|x\|_2 = \Theta(\sqrt{m})\,\|x\|_2 \le b$. However, the exact chance constraint $z_\gamma\|x\|_2 \le b$, where z_γ is the γ-quantile of the univariate standard normal, is independent of the dimension.

Distributionally robust optimization (DRO): DRO sets

$$\hat{\mathscr{F}} = \left\{x \in \mathscr{X} : \inf_{Q\in\mathscr{U}} \mathbb{E}_Q[h(x,\xi)] \ge \gamma\right\}$$

where $\mathscr{U}$ is a set in the space of probability measures that is constructed from data, and is often known as the ambiguity set or uncertainty set. The rationale here is similar to RO, but views the uncertainty in terms of the distribution. If $\mathscr{U}$ is constructed such that it contains the true distribution F with high confidence, i.e., $P_{data}(F \in \mathscr{U}) \ge 1-\beta$, then a solution $\hat{x}^*$ obtained from the DRO will satisfy $P_F(G(\hat{x}^*,\xi) \le b) \ge \gamma$ with at least $1-\beta$ confidence so that (4) holds.

Popular choices of $\mathscr{U}$ include moment sets, i.e., specifying the moments of Q (to be within a range for instance) (Delage and Ye 2010; Wiesemann et al. 2014; Goh and Sim 2010), and distance-based sets, i.e., specifying Q in the neighborhood ball surrounding a baseline distribution, where the ball size is measured by a statistical distance such as ϕ-divergence (Petersen et al. 2000; Ben-Tal et al. 2013; Glasserman and Xu 2014; Lam 2016; Hu and Hong 2013; Jiang and Guan 2016) or Wasserstein distance (Esfahani and Kuhn 2018; Blanchet and Murthy 2019; Gao and Kleywegt 2016).

Ensuring $P_{data}(F \in \mathscr{U}) \ge 1-\beta$ means that $\mathscr{U}$ is a confidence region for F. In the moment set case, this boils down to finding confidence regions for the moments whose sizes in general scale with the probability space dimension. To explain, when only the mean $\mathbb{E}_F[\xi]$ is estimated, the confidence region constructed from the central limit theorem (CLT) takes the form $\{\hat{\mu} + \hat{\Sigma}^{\frac{1}{2}}v : v \in \mathbb{R}^m, \|v\|_2^2 \le \chi^2_{m,1-\beta}\}$, where $\hat{\mu}$ and $\hat{\Sigma}$ are the sample mean and covariance and $\chi^2_{m,1-\beta}$ (which is of order m) is the $1-\beta$ quantile of the χ^2 distribution with degree of freedom m, therefore the diameter of the confidence region scales as $\sqrt{m}$. When the mean

and covariance are jointly estimated, the dimension dependence scales up further. In the distance-based set case, one needs to estimate statistical distances. If the Wasserstein distance is used to construct the ball surrounding the empirical distribution, results from measure concentration (Fournier and Guillin 2015) indicate that the ball size needs to be of order $n^{-\frac{1}{m}}$ to ensure $P_{data}(F \in \mathscr{U}) \geq 1-\beta$. Alternatively, if $\mathscr{U}$ is constructed as a ϕ-divergence ball surrounding some nonparametric kernel-type density estimate, results from kernel density estimation (see Section 4.3 in Wand and Jones 1994) suggests that the estimation error is of order $n^{-\frac{4}{m+4}}$. In either case, the required size of the uncertainty set exhibits exponential dependence on the dimension. Recently, the empirical likelihood method has also been proposed to calibrate the ball size such that $\mathscr{U}$ can be (much) smaller than what is needed in being a confidence region for F, while at the same time (4) still holds (Lam and Zhou 2017;Duchi et al. 2016; Lam 2019; Blanchet and Kang 2016). However, the ball size in this approach scales as the supremum of a so-called χ^2-process over the decision space (e.g., Lam 2019). An analysis using metric entropy (e.g., Example 2 in Section 14 in Lifshits (1995)) shows that the χ^2-process supremum can scale linearly in the decision space dimension d, a much better but still considerable dependence on the dimension.

Finally, we mention that the only exceptional paradigm to our knowledge that provides an alternate guarantee for (3) in the case of CCP, without using (4), is scenario optimization (SO) (e.g., Calafiore and Campi 2005; Campi and Garatti 2008). In its basic form, this approach sets

$$\hat{\mathscr{F}} = \{x \in \mathscr{X} : G(x,\xi_i) \leq b \ \text{ for all } \ i = 1,\ldots,n\}$$

i.e., using sampled constraints formed from the data. As the number of constraints increases, $\hat{\mathscr{F}}$ is postulated to populate the decision space in some sense and ensure the obtained solution $\hat{x}^*$ lies in $\mathscr{F}$. While the sample size required in the basic SO is linear in the decision dimension d, recent works reduce this dependence by using regularization (Campi and Carè 2013), tighter support rank estimates (Schildbach et al. 2013; Campi and Garatti 2018) and validation-type schemes (Carè et al. 2014; Calafiore 2017). The approach that we propose next is closest to some of the validation-type schemes suggested for SO, but substantially more general as it applies beyond CCP and to all the exemplified methods mentioned above.

In the following sections, we will overview our main idea, procedures and guarantees, leaving the full demonstration of our results and mathematical analyses to a journal version of this work. Section 3 explains the rationale of our framework and thereby invokes our two-phase framework. Sections 4 and 5 present two methods to be used in the framework. Section 6 shows a numerical example. Section 7 concludes the paper.

3 FRAMEWORK AND RATIONALE

Our primary goal is to create a framework to obtain solutions from data-driven optimization reformulations that are less conservative than past proposals, especially in terms of the dimension dependence that some of them face. Our framework works for general stochastically constrained problems in (1) including CCP.

Our key observation is the following. In all the described approaches, the data-driven reformulation requires a parameter that controls the level of conservativeness:

1. SAA: safety margin δ
2. RO and SCA: uncertainty set size ρ
3. DRO: neighborhood ball size or moment set size
4. SO: number of constraints

These parameters have the properties that setting it to one extreme (e.g., 0) would signal no uncertainty in the formulation, leading to a solution very likely infeasible, while setting it to another extreme (e.g., ∞) would cover the entire decision space, leading to a very conservative solution. Besides SO, the rationale undertaken in Section 2 is to select these parameters so that the resulting feasible region $\hat{\mathscr{F}}$ is contained

entirely in $\mathscr{F}$, which involves a vast simultaneous estimation problem on the whole decision space and potentially makes the choice of this parameter overly protective.

On the other hand, given a specific data-driven reformulation, it is easy to see that no matter how we choose this "conservativeness" parameter, the solution must lie in a low-dimensional manifold. To be more precise, suppose the given data-driven reformulation is

$$\begin{array}{ll} \min\limits_{x\in\mathscr{X}} & f(x) \\ \text{subject to} & x \in \hat{\mathscr{F}}(s) \end{array} \tag{6}$$

where $s \in \mathbb{R}$ denotes the conservativeness parameter that determines the size of feasible region $\hat{\mathscr{F}}$. We denote the obtained solution from (6) as $x^*(s)$. The *solution path* $\{x^*(s) : s \in \mathbb{R}\}$ contains all possible obtainable solutions from the data-driven reformulation (6). Intuitively, it is only necessary to focus on this solution path, instead of a simultaneous estimation for the whole decision space.

Nonetheless, besides the conservativeness parameter, a data-driven reformulation could have other parameters playing various roles (e.g., center and shape of the set in ellipsoidal RO, baseline distribution in distance-based DRO etc.). The flexibility of these parameter values can enlarge the obtainable solution space and elevate dimension dependence. Suppose we want to contain this enlargement, and at the same time be able to select the optimal candidate among the low-dimensional manifold $\{x^*(s) : s \in \mathbb{R}\}$. We propose the following two-phase framework to achieve this.

Algorithm 1 The Two-Phase Framework

Input: data $\boldsymbol{\xi}_{1:n} = \{\xi_1, \ldots, \xi_n\}$; numbers of data n_1, n_2 allocated to each phase ($n_1 + n_2 = n$); a confidence level $1-\beta$; a given method to construct data-driven reformulation with a (possibly multi-dimensional) parameter $s \in S$; a set of candidate parameter values $\{s_1, s_2, \ldots, s_p\} \subseteq S$.

Phase one:

1. Use n_1 observations, which we index as $\{\xi_{n_2+1}, \ldots, \xi_n\}$ for convenience, to construct the data-driven reformulation $OPT(s)$ in the form (6) parameterized by $s \in S$.

2. For each $j = 1, \ldots, p$, compute the optimal solution $x^*(s_j)$ of $OPT(s_j)$.

Phase two:

Use a validator V to select $(\hat{s}^*, x^*(\hat{s}^*)) = V(\{\xi_1, \ldots, \xi_{n_2}\}, \{x^*(s_1), \ldots, x^*(s_p)\}, 1-\beta)$, where $x^*(\hat{s}^*)$ is a solution and $\hat{s}^*$ is the associated parameter value.

Output: $x^*(\hat{s}^*)$.

Our procedure (Algorithm 1) splits the data into two groups. With the first group of data, we construct a given data-driven reformulation parametrized by a conservativeness parameter s that varies over a space S, which we call $OPT(s)$. We obtain the optimal solution $x^*(s)$ for a range of values $s = s_j, j = 1, \ldots, p$. This step assumes the availability of an efficient solver for $OPT(s)$. Next, the second group of data is fed into a *validator* V that aims to identify the best feasible solution $x^*(\hat{s}^*)$ among $\{x^*(s_j) : j = 1, \ldots, p\}$. The number of points p required to validate depends crucially on the size of S, which is constructed to be low-dimensional. Here for simplicity we assign the last n_1 observations to Phase two and all other to Phase one, and note that it is equivalent to uniformly choose n_1 observations for Phase one because of the i.i.d. nature. There are multiple ways to set up the validator V, each with its own benefits which we will describe precisely in the next two sections. Because of space limits, we solely focus on chance constraints of the form

$$\begin{array}{ll} \min\limits_{x\in\mathscr{X}} & f(x) \\ \text{subject to} & P(x) := P_F((x,\xi) \in A) \geq 1-\alpha \end{array} \tag{7}$$

where A is a deterministic subset of $\mathbb{R}^d \times \mathbb{R}^m$, and $1-\alpha$ is the tolerance level (i.e., $1-\alpha = \gamma$ in (1)).

4 VALIDATION VIA MULTIVARIATE GAUSSIAN SUPREMUM

Our first validator uses a simultaneous estimation of the satisfaction probability $P(x)$ over the discretized solution path of $x^*(s)$. More precisely, given the solution set $\{x^*(s_j) : j = 1,\dots,p\}$, we use a sample average with an appropriately calibrated safety margin, i.e., $(1/n_2)\sum_{i=1}^{n_2}\mathbf{1}((x,\xi_i)\in A)-\varepsilon$, to replace the unknown $P(\cdot)$ in (7) and output the best solution among the set. The margin ε is calibrated via the limiting distribution of $((1/n_2)\sum_{i=1}^{n_2}\mathbf{1}((x^*(s_j),\xi_i)\in A))_{j=1,\dots,p}$ which is multivariate Gaussian. It contains a critical value $q_{1-\beta}$ that is the quantile of a Gaussian supremum. Algorithms 2 and 3 describe two variants of this validator, one unnormalized while another one normalized by the standard deviation at each s_j. In the following, we denote $N_p(0,\Sigma)$ as a p-dimensional Gaussian vector with mean zero and covariance Σ.

Algorithm 2 V: Unnormalized Gaussian Supremum

Input: $\{\xi_1,\dots,\xi_{n_2}\},\{x^*(s_1),\dots,x^*(s_p)\},1-\beta$

1. For each $j=1,\dots,p$ compute the sample mean $\hat{P}_j=(1/n_2)\sum_{i=1}^{n_2}\mathbf{1}((x^*(s_j),\xi_i)\in A)$ and sample covariance matrix $\hat{\Sigma}$ with $\hat{\Sigma}(j_1,j_2)=(1/n_2)\sum_{i=1}^{n_2}(\mathbf{1}((x^*(s_{j_1}),\xi_i)\in A)-\hat{P}_{j_1})(\mathbf{1}((x^*(s_{j_2}),\xi_i)\in A)-\hat{P}_{j_2})$.
2. Compute $q_{1-\beta}$, the $(1-\beta)$-quantile of $\max\{Z_1,\dots,Z_p\}$ where $(Z_1,\dots,Z_p)\sim N_p(0,\hat{\Sigma})$, and let

$$\hat{s}^*=\operatorname{argmin}\Big\{f(x^*(s_j)):\hat{P}_j\geq 1-\alpha+\frac{q_{1-\beta}}{\sqrt{n_2}},1\leq j\leq p\Big\}. \tag{8}$$

Output: $\hat{s}^*,x^*(\hat{s}^*)$.

Algorithm 3 V: Normalized Gaussian Supremum

Input: $\{\xi_1,\dots,\xi_{n_2}\},\{x^*(s_1),\dots,x^*(s_p)\},1-\beta$

1. Same as in Algorithm 2.
2. Denote $\hat{\sigma}_j^2=\hat{\Sigma}(j,j)$. Compute $q_{1-\beta}$, the $(1-\beta)$-quantile of $\max\{Z_j/\hat{\sigma}_j:\hat{\sigma}_j^2>0,1\leq j\leq p\}$ where $(Z_1,\dots,Z_p)\sim N_p(0,\hat{\Sigma})$, and let

$$\hat{s}^*=\operatorname{argmin}\Big\{f(x^*(s_j)):\hat{P}_j\geq 1-\alpha+\frac{q_{1-\beta}\hat{\sigma}_j}{\sqrt{n_2}},1\leq j\leq p\Big\}. \tag{9}$$

Output: $\hat{s}^*,x^*(\hat{s}^*)$.

The first Gaussian supremum validator (Algorithm 2) is reasoned from a joint CLT that governs the convergence of $\sqrt{n_2}(\hat{P}_1-P(x^*(s_1)),\dots,\hat{P}_p-P(x^*(s_p)))$ to $N_p(0,\Sigma)$, where $\Sigma(j_1,j_2)=\mathrm{Cov}(\mathbf{1}((x^*(s_{j_1}),\xi)\in A),\mathbf{1}((x^*(s_{j_2}),\xi)\in A))$. Using the sample covariance $\hat{\Sigma}$ from Step 1 of Algorithm 2 as an approximation of Σ, we have, by the continuous mapping theorem,

$$\max_{1\leq j\leq p}\sqrt{n_2}(\hat{P}_j-P(x^*(s_j)))\approx\max_{1\leq j\leq p}Z_j\ \text{ in distribution}$$

where $(Z_1,\dots,Z_p)\sim N_p(0,\hat{\Sigma})$. Therefore using the $1-\beta$ quantile $q_{1-\beta}$ of the Gaussian supremum in the margin leads to

$$P(x^*(s_j))\geq\hat{P}_j-\frac{q_{1-\beta}}{\sqrt{n_2}}\ \text{ for all }\ j=1,\dots,p,\ \text{ with probability}\approx 1-\beta.$$

The second validator (Algorithm 3) uses an alternate version of the CLT that is normalized by the componentwise standard deviation σ_j, i.e., $\sqrt{n_2}((\hat{P}_1-P(x^*(s_1)))/\sigma_1,\dots,(\hat{P}_p-P(x^*(s_p)))/\sigma_p)$ converges

to $N_p(0, D\Sigma D)$, where D is a diagonal matrix of $1/\sigma_j, j = 1, \ldots, p$. Note that the quantile $q_{1-\beta}$ in both validators can be computed to high accuracy via Monte Carlo.

The following two theorems describe the feasibility guarantees of the obtained solutions output by the validators in Algorithms 2 and 3:

Theorem 1 Let $\bar{\alpha} = 1 - \max_{1\le j\le p} P(x^*(s_j))$. For every solution set $\{x^*(s_j) : 1 \le j \le p\}$, every n_2, and $\beta \in (0, \frac{1}{2})$, the solution output by Algorithm 2 satisfies

$$P_{\boldsymbol{\xi}_{1:n_2}}(x^*(\hat{s}^*) \text{ is feasible for (7)}) \ge 1 - \beta - C\Big(\Big(\frac{\log^7(pn_2)}{n_2\alpha}\Big)^{\frac{1}{6}} + \exp\big(-cn_2 \min\{\varepsilon, \frac{\varepsilon^2}{\bar{\alpha}}\}\big)\Big)$$

with

$$\varepsilon = \Big(\alpha - \bar{\alpha} - C\sqrt{\frac{\log(p/\beta)}{n_2}}\Big)_+ \tag{10}$$

where C and c are universal constants, and $P_{\boldsymbol{\xi}_{1:n_2}}$ denotes the probability with respect to Phase two data $\{\xi_1, \ldots, \xi_{n_2}\}$ conditioned on Phase one data.

Theorem 2 Under the same conditions of Theorem 1, the solution output by Algorithm 3 satisfies

$$P_{\boldsymbol{\xi}_{1:n_2}}(x^*(\hat{s}^*) \text{ is feasible for (7)}) \ge 1 - \beta - C\Big(\Big(\frac{\log^7(pn_2)}{n_2\alpha}\Big)^{\frac{1}{6}} + \frac{\log^2(pn_2)}{\sqrt{n_2\alpha}} + \exp\big(-cn_2 \min\{\varepsilon, \frac{\varepsilon^2}{\bar{\alpha}}\}\big)\Big)$$

with

$$\varepsilon = \Big(\alpha - \bar{\alpha} - C\sqrt{\frac{\big(\bar{\alpha} + \log(n_2\alpha)/n_2\big)\log(p/\beta)}{n_2}}\Big)_+ \tag{11}$$

where C and c are universal constants.

In both Theorems 1 and 2, the finite-sample coverage probability consists of two sources of errors. The first source comes from the CLT approximation that decays polynomially in the Phase 2 sample size n_2. The second error arises from the possibility that none of the solutions $\{x^*(s_1), \ldots, x^*(s_p)\}$ satisfies the criterion in (8) or (9), which vanishes exponentially fast. When ε in (10) and (11) are of constant order, the CLT error dominates. In this case the finite-sample error depends logarithmically on p, the number of candidate parameter values, and the bounds dictate a coverage tending to $1-\beta$ when p is as large as $\exp(o(n_2^{1/7}))$. The derivation of this logarithmic dependence on p utilizes a high-dimensional CLT recently developed in (Chernozhukov et al. 2017).

We explain the implication on the dimensionality of the problem. Note that to sufficiently cover the whole solution path, p is typically exponential in the dimension of S, denoted $\dim(S)$ (this happens when we uniformly discretize the parameter space S). The discussion above thus implies a requirement that n_2 is of higher order than $\dim(S)^7$. Here the low dimensionality of S is crucial; for instance, a one-dimensional conservativeness parameter s would mean $\dim(S) = 1$, so that a reasonably small n_2 can already ensure adequate feasibility coverage. Moreover, the margin adjustments in Algorithms 2 and 3 both depend only on $\dim(S)$. Thus, the choice of $\hat{s}^*$ also depends only on $\dim(S)$, but not the dimension of the whole decision space. This indicates a substantial reduction in conservativeness compared to the approaches described in Section 2.

Comparing between the two validators, we also see that the normalized one (Algorithm 3) is statistically more efficient than the unnormalized one (Algorithm 2) when the tolerance level $1-\alpha$ is large (i.e., the common case). More specifically, in order to make the exponential error non-trivial, one needs at least $\varepsilon > 0$. In the case of Algorithm 2 expression (10) suggests that, after ignoring the logarithmic factor $\log(p/\beta)$, this requires an n_2 of order $(\alpha - \bar{\alpha})^{-2}$, whereas in (11) it can be seen to need only an n_2 of order $\alpha(\alpha - \bar{\alpha})^{-2}$, a much smaller size when $1-\alpha$ is close to 1.

Theorems 1 and 2 also give immediately the following asymptotic feasibility guarantee:

Corollary 1 Let $\bar{\alpha} = 1 - \max_{1\le j\le p} P(x^*(s_j))$. For every solution set $\{x^*(s_j) : 1 \le j \le p\}$ such that $\bar{\alpha} < \alpha$ and every $\beta \in (0, \frac{1}{2})$, the solution output by Algorithm 2 or 3 satisfies

$$\liminf_{n_2\to\infty \text{ and } p\exp(-n_2^{1/7})\to 0} P_{\boldsymbol{\xi}_{1:n_2}}(x^*(\hat{s}^*) \text{ is feasible for (7)}) \ge 1-\beta.$$

5 VALIDATION VIA UNIVARIATE GAUSSIAN MARGIN

We offer an alternate validator that can perform more efficiently than Algorithms 2 and 3, provided that some further assumptions are in place. This is a scheme that simply uses a standard univariate Gaussian critical value to calibrate the margin (Algorithm 4).

Algorithm 4 outputs a solution with objective value no worse than Algorithms 2 and 3. Comparing the criteria to choose $\hat{s}^*$, we see that, thanks to the stochastic dominance of the maximum among a multivariate Gaussian vector over each of its individual components, the margin in (8) satisfies $q_{1-\beta} \ge z_{1-\beta}\hat{\sigma}_j$ for all j, and similarly the margin in (9) satisfies $q_{1-\beta}\hat{\sigma}_j \ge z_{1-\beta}\hat{\sigma}_j$, so that both are bounded from below by the margin of (12). Consequently the solution from (12) achieves an objective value no worse than the other two.

Algorithm 4 V: Univariate Gaussian Validator

Input: $\{\xi_1,\dots,\xi_{n_2}\}, \{x^*(s_1),\dots,x^*(s_p)\}, 1-\beta$

1. For each $j = 1,\dots,p$ compute the sample mean $\hat{P}_j = (1/n_2)\sum_{i=1}^{n_2} \mathbf{1}((x^*(s_j),\xi_i)\in A)$ and sample variance $\hat{\sigma}_j^2 = \hat{P}_j(1-\hat{P}_j)$.

2. Compute

$$\hat{s}^* = \operatorname{argmin}\left\{ f(x^*(s_j)) \middle| \hat{P}_j \ge 1-\alpha + \frac{z_{1-\beta}\hat{\sigma}_j}{\sqrt{n_2}}, 1\le j\le p \right\} \tag{12}$$

where $z_{1-\beta}$ is the $1-\beta$ quantile of the standard Gaussian distribution.

Output: $\hat{s}^*, x^*(\hat{s}^*)$.

The validity of the univariate Gaussian critical value is based on the statistical consistency of the obtained solution $x^*(\hat{s}^*)$ to some limiting solution (correspondingly $\hat{s}^*$ to some limiting optimal parameter value) as n_2 increases. Intuitively, this implies that with sufficient sample size one can focus feasibility validation on a small neighborhood of $\hat{s}^*$, which further suggests that we need to control only the statistical error at effectively one solution parametrized at $\hat{s}^*$. For this argument to hold, however, we would need additional assumptions on the CCP (7):

Assumption 1 (Continuous objective) The objective function $f(x)$ is continuous on $\mathscr{X}$.

Assumption 2 (Linear constraint) $\mathbf{1}((x,\xi)\in A) = \mathbf{1}(a_k^T x \le b_k \text{ for } k = 1,\dots,K)$, where each a_k has a density on $\mathbb{R}^d$ and each b_k is a non-zero constant.

We comment that Assumption 2 is a sufficient but not necessary condition, and in fact some of the results presented below holds for more general constraints. For instance, results in Theorem 3 remain valid for constraints of the form $\mathbf{1}(a_k^T A_k(x) \le b_k \text{ for } k = 1,\dots,K)$ where each $A_k(\cdot) : \mathbb{R}^d \to \mathbb{R}^m$ is a continuous function. We further assume a smoothness condition for the data-driven reformulation $OPT(s), s \in S$:

Assumption 3 (Piecewise continuous solution curve) The parameter space S is a finite interval $[s_l, s_u]$. The optimal solution $x^*(s)$ of $OPT(s)$ exists and is unique except for a finite number of parameter values $\tilde{s}_i, i = 1,\dots,M-1$ such that $s_l = \tilde{s}_0 < \tilde{s}_1 < \cdots < \tilde{s}_{M-1} < \tilde{s}_M = s_u$, and the parameter-solution mapping $x^*(s)$ is uniformly continuous on $[\tilde{s}_0, \tilde{s}_1)$, $(\tilde{s}_{M-1}, \tilde{s}_M]$, and $(\tilde{s}_{i-1}, \tilde{s}_i)$ for $i = 2,\dots,M-1$.

Continuity of the solution path allows approximating the whole solution curve by discretizing the parameter space S. Also note that under Assumption 3 the solution $x^*(s)$ exists and is unique for almost surely every $s \in S$ with respect to the Lebesgue measure. Therefore, if one discretizes the parameter

space by randomizing via a continuous distribution over S, then with probability one the solution $x^*(s)$ is unique at all sampled parameter values. This provides an easy way to ensure the assumption that none of the parameter values $\{s_1,\ldots,s_m\}$ used in Phase one of Algorithm 1 belongs to the discontinuity set $\{\tilde{s}_1,\ldots,\tilde{s}_{M-1}\}$.

To explain the superior performance of Algorithm 4, we introduce a notion of optimality within the solution path $\{x^*(s): s\in S\}$. First we fill in the holes of of the solution curve. Under piecewise uniform continuity, the parameter-solution mapping $x^*(s)$ on each piece $(\tilde{s}_{i-1},\tilde{s}_i)$ can be continuously extended to the closure $[\tilde{s}_{i-1},\tilde{s}_i]$. Specifically, we define the extended parameter-solution mapping at the endpoints of the i-th piece

$$x_i^*(\tilde{s}_{i-1}) := \lim_{s\to\tilde{s}_{i-1}+} x^*(s),\ x_i^*(\tilde{s}_i) := \lim_{s\to\tilde{s}_i-} x^*(s)$$

and $x_i^*(s)=x^*(s)$ for $s\in(\tilde{s}_{i-1},\tilde{s}_i)$. $x_{i+1}^*(\tilde{s}_i)$ and $x_i^*(\tilde{s}_i)$ take different values if the i-th and $(i+1)$-th pieces are disconnected. With the extended parameter-solution mappings $x_i^*(\cdot)$'s, we define the following as the optimal solution set among the solution set parameterized by s

$$\mathscr{X}_S^* := \operatorname{argmin}\{f(x): P(x)\geq 1-\alpha, x=x_i^*(s) \text{ for some } s\in[\tilde{s}_{i-1},\tilde{s}_i] \text{ and } i=1,\ldots,M\}$$

and the define the set of optimal parameter as

$$S^* := \{s: s\in[\tilde{s}_{i-1},\tilde{s}_i] \text{ and } x_i^*(s)\in\mathscr{X}_S^* \text{ for some } i=1,\ldots,M\}.$$

We need several additional technical assumptions. The first is that the chance constraint is not binding at the endpoints of each piece of the solution path:

Assumption 4 $P(x_i^*(\tilde{s}_i))\neq 1-\alpha$ and $P(x_{i+1}^*(\tilde{s}_i))\neq 1-\alpha$ for all $i=1,\ldots,M-1$, $P(x^*(s_l))\neq 1-\alpha$, $P(x^*(s_u))\neq 1-\alpha$, and $\sup_{s\notin\{\tilde{s}_1,\ldots,\tilde{s}_{M-1}\}} P(x^*(s))>1-\alpha$.

Since the solution path $\{x^*(s): s\in S\}$ depends on the Phase one data $\boldsymbol{\xi}_{n_2+1:n}$, the path and hence the endpoints $x_i^*(\tilde{s}_i), x_{i+1}^*(\tilde{s}_i)$ are random objects, and so the first part of Assumption 4 is expected to hold almost surely provided that the set $\{x\in\mathscr{X}: P(x)=1-\alpha\}$ is a null set under the Lebesgue measure. The second part states that the solution path contains a strictly feasible solution which in turn ensures that the optimal solution set $\mathscr{X}_S^*$ is non-empty. Note that this can be achieved by simply including very conservative parameter values in S.

Another property we assume regards the monotonicity of the feasible set size with respect to the parameter s in the reformulation $OPT(s)$:

Assumption 5 Denote by $\mathrm{Sol}(s)\subseteq\mathscr{X}$ the feasible set of $OPT(s)$. Assume $\mathrm{Sol}(s)$ is a closed set for all $s\in S$ and $\mathrm{Sol}(s_2)\subseteq\mathrm{Sol}(s_1)$ for all $s_1,s_2\in S$ such that $s_1<s_2$.

Assmption 5 holds for almost all common reformulations because of a monotonic relation between the parameter s and the conservativeness level. For instance, in RO with ellipsoidal uncertainty set, the RO feasible region shrinks with the radius of the ellipsoid, and similar relations hold for DRO, SAA, and SO.

Lastly, we assume the following technical assumption for the set of optima:

Assumption 6 For any $\varepsilon>0$ there exists an $s\notin\{\tilde{s}_1,\ldots,\tilde{s}_{M-1}\}$ such that $P(x^*(s))>1-\alpha$ and $d(x^*(s),\mathscr{X}_S^*)<\varepsilon$, where $d(x^*(s),\mathscr{X}_S^*) := \inf_{x\in\mathscr{X}_S^*}\|x^*(s)-x\|_2$.

This assumption trivially holds if we have $\max_{x\in\mathscr{X}_S^*} P(x)>1-\alpha$. Otherwise, if $P(x)=1-\alpha$ for all $x\in\mathscr{X}_S^*$, it rules out the case that the solution path $x^*(s)$ passes through the optima without entering the interior of the feasible set of (7). This extreme case typically happens with zero probability, again in view of the fact that the solution path is itself random with respect to Phase one data.

With these assumptions, we have the following asymptotic performance guarantee for Algorithm 4:

Theorem 3 Under Assumptions 1-6 hold, the optimal solution set is a singleton, i.e., $\mathscr{X}_S^*=\{x_S^*\}$. Denote by $\varepsilon_s=\sup_{s\in S}\inf_{1\leq j\leq p}|s-s_j|$ the mesh size. We also have that, with respect to $\{\xi_1,\ldots,\xi_{n_2}\}$, the solution

and parameter output by Algorithm 4 satisfy

$$\lim_{n_2\to\infty,\varepsilon_s\to 0} x^*(\hat{s}^*) = x_S^* \text{ and } \lim_{n_2\to\infty,\varepsilon_s\to 0} d(\hat{s}^*, S^*) = 0$$

almost surely, where $d(\hat{s}^*, S^*) = \inf_{s\in S^*} |\hat{s}^* - s|$. Moreover

$$\begin{cases} \liminf_{n_2\to\infty,\varepsilon_s\to 0} P_{\boldsymbol{\xi}_{1:n_2}}(x^*(\hat{s}^*) \text{ is feasible for (7)}) \geq 1-\beta & \text{if } P(x_S^*) = 1-\alpha \\ \lim_{n_2\to\infty,\varepsilon_s\to 0} P_{\boldsymbol{\xi}_{1:n_2}}(x^*(\hat{s}^*) \text{ is feasible for (7)}) = 1 & \text{if } P(x_S^*) > 1-\alpha. \end{cases}$$

Theorem 3 states that as the mesh $\{s_1,\ldots,s_p\}$ gets increasingly fine and and the data size grows, the solution given by Algorithm 4 enjoys performance guarantees concerning both feasibility and optimality. In particular, the estimated solution and the parameter converge to the unique optimal solution x_S^* and the optimal parameter set S^* respectively, and simultaneously the obtained solution $x^*(\hat{s}^*)$ is feasible with the desired confidence level $1-\beta$.

6 A NUMERICAL EXAMPLE

We present a numerical example to demonstrate the performances of our framework. We consider the following linear CCP

$$\min\ c^T x \ \text{ subject to } \ P_F(\xi^T x \leq b) \geq 1-\alpha$$

where $c \in \mathbb{R}^d, b \in \mathbb{R}$ are deterministic, the distribution F of the randomness $\xi \in \mathbb{R}^d$ is multivariate Gaussian with mean μ and covariance Σ, and the tolerance level $1-\alpha$ is set to 90%.

We test the proposed framework on RO with ellipsoid uncertainty set that leads to a robust counterpart in the form $\hat{\mu}^T x + \sqrt{s}\|\hat{\Sigma}^{1/2}x\|_2 \leq b$ where $\hat{\mu}$ and $\hat{\Sigma}$ are the sample mean and covariance for ξ computed from Phase one data. The benchmark ("SCA" in the table) is set to an SCA (equation 2.4.11 of Ben-Tal et al. 2009), which in our case can be expressed as $\mu' x + \sqrt{2\log(1/\alpha)}\|\Sigma^{1/2}x\|_2 \leq b$. Here, we give this SCA or RO the advantage of knowing the true mean μ and covariance Σ of the randomness. Note that this is equivalent to setting $s = 2\log\frac{1}{\alpha}$. To implement our framework, we take the $(1-\alpha)n_1$-th order statistic $\hat{s}_{1-\alpha}$ of $\{(\xi_{n_2+i}-\hat{\mu})'\hat{\Sigma}^{-1}(\xi_{n_2+i}-\hat{\mu}) : i = 1,\ldots,n_1\}$, where $\xi_{n_2+i}, i = 1,\ldots,n_1$ are the Phase one data, so that $\{\xi : (\xi-\hat{\mu})'\hat{\Sigma}^{-1}(\xi-\hat{\mu}) \leq \hat{s}_{1-\alpha}\}$ is roughly a $(1-\alpha)$-content set for ξ (such type of quantile-based selection has been used in Hong et al. 2017). We then set the values $s_j = (\hat{s}_{1-\alpha}+20)j/50$ for $j = 1,\ldots,50$.

Table 1 summarizes the results under dimension $d = 10$ and data sizes $n = 200, 500$. "unnorm. GS", "norm. GS" and "uni. Gaussian" denote Algorithms 2, 3 and 4 respectively. For each setting we repeat the experiments 1000 times each with an independently generated data set and a data-driven solution output, and then take down the average objective value ("mean obj. val.") achieved by these solutions and the proportion ("feasibility level") of feasible solutions as the empirical feasibility coverage. Therefore, the smaller the "mean obj. val." is, the better is the solution in terms of optimality, and "feasibility level" $\geq 95\%$ indicates that the desired feasibility confidence level 95% is achieved and otherwise not.

Table 1: RO with ellipsoidal uncertainty set. $d = 10$. Data are split to $n_1 = n_2 = n/2$.

	SCA	unnorm. GS		norm. GS		uni. Gaussian	
		$n = 200$	$n = 500$	$n = 200$	$n = 500$	$n = 200$	$n = 500$
mean obj. val.	-3.57	-3.68	-4.42	-4.20	-4.58	-4.43	-4.80
feasibility level	100%	99.9%	99.8%	98.5%	99.6%	97.5%	98.8%

We highlight two observations. First, our framework with the three proposed validators outperforms the SCA benchmark. In terms of the objective performance, all our validators achieve lower objective value than SCA (with a difference ≥ 0.6), while at the same time retain the feasibility confidence to above 95% in all the three tables. Second, among the three proposed validators, the univariate Gaussian validator

appears less conservative than the Gaussian supremum counterparts in achieving better objective values, and relatedly tighter feasibility confidence levels (i.e., closer to 95%).

7 CONCLUSION

We have studied a validation-based framework to reduce the conservativeness in data-driven optimization with uncertain constraints, faced by the several conventional approaches that rely on implicit estimation of the entire feasible region. Our framework leverages the low dimensionality of the possible solutions output from a data-driven reformulation class, by extracting the associated parametrized solution path and selecting the best candidate within the path. We have proposed several validators for solution selection, and have presented results on the feasibility guarantees on the obtained solutions under our validators that scale favorably with the problem dimension. We have illustrated the benefits of our approach with a numerical example.

ACKNOWLEDGMENTS

We gratefully acknowledge support from the National Science Foundation under grants CAREER CMMI-1653339/1834710 and IIS-1849280.

REFERENCES

Atlason, J., M. A. Epelman, and S. G. Henderson. 2004. "Call Center Staffing with Simulation and Cutting Plane Methods". *Annals of Operations Research* 127(1-4):333–358.

Ben-Tal, A., D. Den Hertog, A. De Waegenaere, B. Melenberg, and G. Rennen. 2013. "Robust Solutions of Optimization Problems Affected by Uncertain Probabilities". *Management Science* 59(2):341–357.

Ben-Tal, A., L. El Ghaoui, and A. Nemirovski. 2009. *Robust Optimization*. Princeton, New Jersey: Princeton University Press.

Bertsimas, D., D. B. Brown, and C. Caramanis. 2011. "Theory and Applications of Robust Optimization". *SIAM Review* 53(3):464–501.

Bertsimas, D., V. Gupta, and N. Kallus. 2018. "Data-Driven Robust Optimization". *Mathematical Programming* 167(2):235–292.

Blanchet, J., and Y. Kang. 2016. "Sample Out-Of-Sample Inference Based on Wasserstein Distance". *arXiv preprint arXiv:1605.01340*.

Blanchet, J., and K. Murthy. 2019. "Quantifying Distributional Model Risk via Optimal Transport". *Mathematics of Operations Research* 44(2):565–600.

Calafiore, G., and M. C. Campi. 2005. "Uncertain Convex Programs: Randomized Solutions and Confidence Levels". *Mathematical Programming* 102(1):25–46.

Calafiore, G. C. 2017. "Repetitive Scenario Design". *IEEE Transactions on Automatic Control* 62(3):1125–1137.

Campi, M. C., and A. Carè. 2013. "Random Convex Programs with L_1-Regularization: Sparsity and Generalization". *SIAM Journal on Control and Optimization* 51(5):3532–3557.

Campi, M. C., and S. Garatti. 2008. "The Exact Feasibility of Randomized Solutions of Uncertain Convex Programs". *SIAM Journal on Optimization* 19(3):1211–1230.

Campi, M. C., and S. Garatti. 2018. "Wait-And-Judge Scenario Optimization". *Mathematical Programming* 167(1):155–189.

Carè, A., S. Garatti, and M. C. Campi. 2014. "Fast—Fast Algorithm for the Scenario Technique". *Operations Research* 62(3):662–671.

Chernozhukov, V., D. Chetverikov, and K. Kato. 2017. "Central Limit Theorems and Bootstrap in High Dimensions". *The Annals of Probability* 45(4):2309–2352.

Delage, E., and Y. Ye. 2010. "Distributionally Robust Optimization Under Moment Uncertainty with Application to Data-Driven Problems". *Operations Research* 58(3):595–612.

Duchi, J., P. Glynn, and H. Namkoong. 2016. "Statistics of Robust Optimization: A Generalized Empirical Likelihood Approach". *arXiv preprint arXiv:1610.03425*.

Esfahani, P. M., and D. Kuhn. 2018. "Data-Driven Distributionally Robust Optimization Using the Wasserstein Metric: Performance Guarantees and Tractable Reformulations". *Mathematical Programming* 171(1-2):115–166.

Fournier, N., and A. Guillin. 2015. "On the Rate of Convergence in Wasserstein Distance of the Empirical Measure". *Probability Theory and Related Fields* 162(3-4):707–738.

Gao, R., and A. J. Kleywegt. 2016. "Distributionally Robust Stochastic Optimization with Wasserstein Distance". *arXiv preprint arXiv:1604.02199*.

Glasserman, P., and X. Xu. 2014. "Robust Risk Measurement and Model Risk". *Quantitative Finance* 14(1):29–58.

Goh, J., and M. Sim. 2010. "Distributionally Robust Optimization and Its Tractable Approximations". *Operations Research* 58(4-part-1):902–917.

Goldfarb, D., and G. Iyengar. 2003. "Robust Portfolio Selection Problems". *Mathematics of Operations Research* 28(1):1–38.

Hong, L. J., Z. Huang, and H. Lam. 2017. "Learning-Based Robust Optimization: Procedures and Statistical Guarantees". *arXiv preprint arXiv:1704.04342*.

Hu, Z., and L. J. Hong. 2013. "Kullback-Leibler Divergence Constrained Distributionally Robust Optimization". *Available at Optimization Online*.

Jiang, R., and Y. Guan. 2016. "Data-Driven Chance Constrained Stochastic Program". *Mathematical Programming* 158(1-2):291–327.

Krokhmal, P., J. Palmquist, and S. Uryasev. 2002. "Portfolio Optimization with Conditional Value-At-Risk Objective and Constraints". *Journal of Risk* 4(2):43–68.

Lam, H. 2016. "Robust Sensitivity Analysis for Stochastic Systems". *Mathematics of Operations Research* 41(4):1248–1275.

Lam, H. 2019. "Recovering Best Statistical Guarantees via the Empirical Divergence-Based Distributionally Robust Optimization". *Operations Research* 67(4):1090–1105.

Lam, H., and E. Zhou. 2017. "The Empirical Likelihood Approach to Quantifying Uncertainty in Sample Average Approximation". *Operations Research Letters* 45(4):301–307.

Lifshits, M. A. 1995. *Gaussian Random Functions*, Volume 322 of *Mathematics and Its Applications*. Dordrecht, Netherlands: Kluwer Academic Publisher.

Luedtke, J., and S. Ahmed. 2008. "A Sample Approximation Approach for Optimization with Probabilistic Constraints". *SIAM Journal on Optimization* 19(2):674–699.

Nemirovski, A. 2003. "On Tractable Approximations of Randomly Perturbed Convex Constraints". In 42nd *IEEE International Conference on Decision and Control*, Volume 3, 2419–2422. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Nemirovski, A., and A. Shapiro. 2006. "Convex Approximations of Chance Constrained Programs". *SIAM Journal on Optimization* 17(4):969–996.

Petersen, I. R., M. R. James, and P. Dupuis. 2000. "Minimax Optimal Control of Stochastic Uncertain Systems with Relative Entropy Constraints". *IEEE Transactions on Automatic Control* 45(3):398–412.

Prékopa, A. 2003. "Probabilistic Programming". In *Stochastic Programming*, Volume 10 of *Handbooks in Operations Research and Management Science*, 267–351. New York: Elsevier.

Schildbach, G., L. Fagiano, and M. Morari. 2013. "Randomized Solutions to Convex Programs with Multiple Chance Constraints". *SIAM Journal on Optimization* 23(4):2479–2501.

Tulabandhula, T., and C. Rudin. 2014. "Robust Optimization Using Machine Learning for Uncertainty Sets". *arXiv preprint arXiv:1407.1097*.

Wand, M. P., and M. C. Jones. 1994. *Kernel Smoothing*. London, UK: Chapman and Hall/CRC.

Wang, W., and S. Ahmed. 2008. "Sample Average Approximation of Expected Value Constrained Stochastic Programs". *Operations Research Letters* 36(5):515–519.

Wiesemann, W., D. Kuhn, and M. Sim. 2014. "Distributionally Robust Convex Optimization". *Operations Research* 62(6):1358–1376.

AUTHOR BIOGRAPHIES

HENRY LAM is an Associate Professor in the Department of Industrial Engineering and Operations Research at Columbia University. His research focuses on Monte Carlo simulation, uncertainty quantification, risk analysis, and stochastic and robust optimization. His email address is khl2114@columbia.edu.

HUAJIE QIAN is a Ph.D. student in the Department of Industrial Engineering and Operations Research at Columbia University. His research interest lies in simulation uncertainty quantification, data-driven simulation analysis, and stochastic optimization. His email address is hq2157@columbia.edu.