Proceedings of the 2016 Winter Simulation Conference
T. M. K. Roeder, P. I. Frazier, R. Szechtman, and E. Zhou, eds.

APPROXIMATING DATA-DRIVEN JOINT CHANCE-CONSTRAINED PROGRAMS VIA UNCERTAINTY SET CONSTRUCTION

L. Jeff Hong

Department of Management Sciences
City University of Hong Kong
86 Tat Chee Ave.
Kowloon Tong, Hong Kong

Zhiyuan Huang, Henry Lam

Department of Industrial and Operations Engineering
University of Michigan
1205 Beal Ave.
Ann Arbor, MI 48109, USA

ABSTRACT

We study the use of robust optimization (RO) in approximating joint chance-constrained programs (CCP), in situations where only limited data, or Monte Carlo samples, are available in inferring the underlying probability distributions. We introduce a procedure to construct uncertainty set in the RO problem that translates into provable statistical guarantees for the joint CCP. This procedure relies on learning the high probability region of the data and controlling the region's size via a reformulation as quantile estimation. We show some encouraging numerical results.

1 INTRODUCTION

Consider a joint chance-constrained program (CCP)

$$\begin{array}{ll} \min & f(x) \\ \text{subject to} & P(g(x;\xi) \in \mathscr{A}) \geq 1-\alpha \end{array} \tag{1}$$

where $x \in \mathbb{R}^m$ is the decision variable, $\xi \in \mathbb{R}^d$ is some random variable under the probability measure P, and $g(x;\xi) : \mathbb{R}^m \times \mathbb{R}^d \to \Omega$ with $\mathscr{A} \in \Omega$. CCPs arise in many different application areas, e.g., reservoir system design (Prékopa 1978), cash matching (Dentcheva 2004) and wireless cooperative network (Shi et al. 2015). They were first proposed by Charnes et al. (1958), Miller and Wagner (1965) and Prékopa (1970), and have been studied extensively in the stochastic programming literature (see Prékopa (2003) for a thorough introduction).

We are primarily interested in situations where P is unknown but limited data are available. As known in the CCP literature (Prékipa 2003), directly solving (1) is difficult unless in few restrictive conditions. In the case of limited data (or Monte Carlo samples), the problem becomes even more challenging. Existing approach includes scenario generation (SG) (Calafiore and Campi 2005, 2006), mixed integer programming (MIP) (Luedtke et al. 2007), and sequential convex approximation (SCA) (Hong et al. 2011). The SG approach requires a specific sample size to satisfy the chance constraint and does not provide the flexibility if the sample size differs from the number. The MIP approach is computationally challenging and it can hardly solve problems with more than a few hundred samples. The SCA approach converges to a KKT point of (1) only asymptotically and it does not provide any finite-sample statistical guarantee.

In this paper, we explore the use of robust optimization (RO) to approximate (1), namely

$$\begin{array}{ll} \min & f(x) \\ \text{subject to} & g(x;\xi) \in \mathscr{A} \quad \text{for all} \quad \xi \in \mathscr{U} \end{array} \tag{2}$$

By choosing $\mathscr{U}$ such that $P(\xi \in \mathscr{U}) \geq 1-\alpha$, the feasible region of (2) is subsisted by (1), and hence the optimal solution from (2) is feasible for (1). Moreover, with a suitably selected shape of $\mathscr{U}$, (2) can be

reformulated into a tractable optimization program, and thus provides a tractable upper approximation for (1).

The idea of using RO to approximate CCP has been suggested in the RO literature (see, for instance, the monograph of Ben-Tal et al. (2009)). In this literature, the size of the uncertainty set $\mathscr{U}$ is commonly calibrated via the use of concentration bounds that take into account information of the random object ξ (e.g., support, mean, variance etc.), which is often loose in practice. Our goal is to study a new approach for constructing $\mathscr{U}$ in a way that can ensure a prescribed level of statistical guarantees, and moreover follows closely the geometry of the data, thus resulting in tight upper approximation. As our analysis reveals, our approach tends to require much less samples than SG in achieving a similar level of statistical accuracy, and shows competitive numerical performance regarding optimality.

2 TWO-PHASE PROCEDURE IN CONSTRUCTING UNCERTAINTY SETS

We introduce our procedure to construct $\mathscr{U}$. We assume a given i.i.d. data set $D = \{\xi_1, \ldots, \xi_n\}$, where $\xi_i \in \mathbb{R}^d$ are sampled under P. Our basic requirement is that $\mathscr{U} = \mathscr{U}(D)$ is a $1-\alpha$ prediction set for P, with a prescribed level of confidence $1-\beta$. In other words,

$$P_D(P(\xi \in \mathscr{U}(D)) \geq 1-\alpha)) \geq 1-\beta \tag{3}$$

where $P_D(\cdot)$ is the probability taken with respect to the data D. This basic requirement clearly translates to the implication that the optimal solution of (2) is feasible for (1) with confidence $1-\beta$.

Note that (3) only focuses on feasibility guarantee for (1), and the resulting solution, even though feasible, can be very conservative. This issue can be alleviated via judiciously choosing $\mathscr{U}$ in two aspects: 1) We prefer $\mathscr{U}$ such that $P(\xi \in \mathscr{U}(D))$ is close to, not just larger than, $1-\alpha$ with high confidence. 2) We prefer $\mathscr{U}$ that has smaller volume. Note that this leads to a larger feasible region in (2), which in turn results in a less conservative approximation to (1). Also note that, with a fixed α, a small $\mathscr{U}$ means a $\mathscr{U}$ that contains a high probability region of ξ.

The above discussion motivates us to propose a two-phase strategy in constructing $\mathscr{U}$. We first split the data D into two groups, denoted D_1 and D_2, with sizes n_1 and n_2 respectively. Say $D_1 = \{\xi_1^1, \ldots, \xi_{n_1}^1\}$ and $D_2 = \{\xi_1^2, \ldots, \xi_{n_2}^2\}$. These two data groups are used in the following two phases:

Phase 1 We use D_1 to calibrate the shape of the high probability region with tractable geometric tools. Two common choices are:

1. *Ellipsoid:* Take the sample mean μ of D_1 and some covariance matrix Σ, e.g., the sample covariance matrix, diagonalized covariance matrix, or identity matrix. Set the shape as $(\xi-\mu)'\Sigma^{-1}(\xi-\mu) \leq \rho$ for some $\rho > 0$.
2. *Polytope:* Form a convex hull of the data in D_1, or alternately, a convex hull of the data that leaves out $\lfloor n_1\alpha \rfloor$ of D_1 that are in the "periphery", e.g. having the smallest Tukey depth (e.g., Serfling (2002) and Hallin et al. (2010)). This results in a polytope represented by $\mathscr{S} = \{\xi : a_i'\xi \leq b_i, i = 1, \ldots, m\}$ where $a_i \in \mathbb{R}^d$ and $b_i \in \mathbb{R}$.

Phase 2 We use D_2 to calibrate the size of the uncertainty set so that it satisfies (3) and moreover $P(\xi \in \mathscr{U}(D)) \approx 1-\alpha$ with high confidence. To do so, we first express our geometric shape calibrated in Phase 1 in the form $\{\xi : t(\xi) \leq s\}$, where $t(\cdot) : \mathbb{R}^d \to \mathbb{R}$ is a transformation mapping from the space of ξ to $\mathbb{R}$, and $s \in \mathbb{R}$. For the two geometric shapes we have considered above,

1. *Ellipsoid:* We set $t(\xi) = (\xi-\mu)'\Sigma^{-1}(\xi-\mu)$.
2. *Polytope:* Find a point (e.g. the Chebyshev center (Boyd and Vandenberghe 2004)), say μ, in $\mathscr{S}^\circ$, the interior of $\mathscr{S}$. Let

$$t(\xi) = \max_{i=1,\ldots,m} \frac{a_i'(\xi-\mu)}{b_i - a_i'\mu}$$

which is well-defined since $\mu \in \mathscr{S}^\circ$. Note that $\{\xi : t(\xi) \le 1\}$ is equivalent to $\mathscr{S}$ defined above.

Our key step is to choose the value of s in the representation $\{\xi : t(\xi) \le s\}$. We use the following lemma:

Lemma 1 Let $Y_1, \ldots, Y_n$ be i.i.d. continuous data in $\mathbb{R}$. Let $Y_{(1)} < Y_{(2)} < \cdots < Y_{(n)}$ be the ordered statistics. A $1-\beta$ confidence upper bound for the γ-quantile of the underlying distribution is $Y_{(r^*)}$, where

$$r^* = \min\left\{ r : \sum_{k=0}^{r-1} \binom{n}{k} \gamma^k (1-\gamma)^{n-k} \ge 1-\beta \right\}$$

If $\sum_{k=0}^{r-1} \binom{n}{k} \gamma^k (1-\gamma)^{n-k} < 1-\beta$ for all $r = 1, \ldots, n$, then none of the $Y_{(r)}$'s is a valid confidence upper bound.

Similarly, a $1-\beta$ confidence lower bound for the γ-quantile of the underlying distribution is $Y_{(r_*)}$, where

$$r_* = \max\left\{ r : \sum_{k=r}^{n} \binom{n}{k} \gamma^k (1-\gamma)^{n-k} \ge 1-\beta \right\}$$

If $\sum_{k=r}^{n} \binom{n}{k} \gamma^k (1-\gamma)^{n-k} < 1-\beta$ for all $r = 1, \ldots, n$, then none of the $Y_{(r)}$'s is a valid confidence lower bound.

Proof. Let q_γ be the γ-quantile. Consider

$$\begin{aligned} P(Y_{(r)} \ge q_\gamma) &= P(\le r-1 \text{ of the data } \{Y_1, \ldots, Y_n\} \text{ are } < q_\gamma) \\ &= \sum_{k=0}^{r-1} \binom{n}{k} F(q_\gamma)^k \bar{F}(q_\gamma)^{n-k} \\ &= \sum_{k=0}^{r-1} \binom{n}{k} \gamma^k (1-\gamma)^{n-k} \end{aligned}$$

by the definition of q_γ. Hence any r such that $\sum_{k=0}^{r-1} \binom{n}{k} \gamma^k (1-\gamma)^{n-k} \ge 1-\beta$ is a $1-\beta$ confidence upper bound for q_γ, and we pick the smallest one. Note that if $\sum_{k=0}^{r-1} \binom{n}{k} \gamma^k (1-\gamma)^{n-k} < 1-\beta$ for all r, then none of the $Y_{(r)}$ is a valid confidence upper bound.

Similarly, we have

$$\begin{aligned} P(Y_{(r)} \le x_p) &= P(\ge r \text{ of the data } \{Y_1, \ldots, Y_n\} \text{ are } \le q_\gamma) \\ &= \sum_{k=r}^{n} \binom{n}{k} F(q_\gamma)^k \bar{F}(q_\gamma)^{n-k} \\ &= \sum_{k=r}^{n} \binom{n}{k} \gamma^k (1-\gamma)^{n-k} \end{aligned}$$

by the definition of q_γ. Hence any r such that $\sum_{k=r}^{n} \binom{n}{k} \gamma^k (1-\gamma)^{n-k} \ge 1-\beta$ will be a $1-\beta$ confidence lower bound for q_γ, and we pick the largest one. Note that if $\sum_{k=r}^{n} \binom{n}{k} \gamma^k (1-\gamma)^{n-k} < 1-\beta$ for all r, then none of the $Y_{(r)}$ is a valid confidence lower bound.

□

Using Lemma 1, with $t(\cdot)$ defined using D_1, we consider the one-dimensional transformed Phase 2 data $t(\xi_1^2), \ldots, t(\xi_{n_2}^2)$. Order them as $t(\xi_{(1)}^2) < \cdots < t(\xi_{(n_2)}^2)$. Set $s = t(\xi_{(r^*)}^2)$, where

$$r^* = \min\left\{ r : \sum_{k=0}^{r-1} \binom{n}{k} (1-\alpha)^k \alpha^{n-k} \ge 1-\beta \right\}$$

The set $\mathscr{U} = \{\xi : t(\xi) \le s\}$ now satisfies (3).

Note that the choice of s above is the smallest among all ordered statistics $t(\xi^2_{(r)})$ that can provide the guarantee (3). It makes $P_D(P(\xi \in \mathscr{U}) \approx 1-\alpha) \ge 1-\beta$ when n_2 is large. For any finite sample, this choice of s can be complemented by the number $t(\xi^2_{(r_*)})$ with

$$r_* = \min\left\{ r : \sum_{k=r}^{n} \binom{n}{k} (1-\alpha)^k \alpha^{n-k} \ge 1-\beta \right\}$$

which, according to Lemma 1 again, provides a lower bound on the size of the uncertainty set in covering $1-\alpha$ portion of ξ with confidence $1-\beta$.

3 VISUALIZATION OF UNCERTAINTY SETS

We have mentioned two choices of uncertainty sets, ellipsoidal and polyhedral. This section visualizes their coverage on the data and also shows some numerical performance. Note that the shape of the uncertainty set does not affect the statistical validity enforced by (3). However, choosing an appropriate shape provides a better solution in terms of optimization. Moreover, it could avoid some computation issues, such as encountering singular covariance matrices.

Throughout this section we consider ξ generated from a bivariate normal distribution with mean $\mu = [1,2]$ and covariance $\Sigma = [2.5, 0.5; 0.5, 1]$. We use a sample size $N = 300$, and use $n_1 = 150$ and $n_2 = 150$ for Phases I and II. The chance constraint parameter and the confidence level is $\alpha = 0.05$ and $\beta = 0.05$ respectively.

For each experiment that corresponds to a specific geometric shape, we generate $1,000$ different data sets of size N, and uses our uncertainty set generation procedure to get $1,000$ $\mathscr{U}$'s. To assess the coverage of each $\mathscr{U}$, we independently generate $10,000$ data points from the same distribution to estimate $\tilde{\alpha} = \hat{P}(\xi \notin \mathscr{U})$. We then estimate the mean set prediction probability $\hat{\alpha}$ as the average of all $1,000$ $\tilde{\alpha}$'s, and $\hat{\beta} = \hat{P}(\tilde{\alpha} > \alpha)$ as the number of times from the $1,000$ replications where $\tilde{\alpha} > \alpha$. We mention that the quantity $P_D(P(\xi \in \mathscr{U}(D)) \ge 1-\alpha)$ guaranteed by our procedure, which we denote as $\beta_{theoretical}$, is actually explicitly known and is equal to 0.0182 for $\alpha = 0.05$ and $n_2 = 150$. This calculation will be elaborated in the next section, and the bottom line here is that the uncertainty set we construct should have $\hat{\beta}$ approximately equal to 0.0182.

3.1 Ellipsoidal Uncertainty Set

3.1.1 General Ellipsoidal Uncertainty Set

A general ellipsoidal set $(\xi - \mu)'\Sigma^{-1}(\xi - \mu)$ is constructed by putting μ as the sample mean and Σ as the sample covariance of D_1. Figure 1 shows a general ellipsoidal set constructed from our two-phase procedure applied on one realization of 300 data points. Our set seems to cover the high density region of the data very well, which is not surprising since the data set comes from a normal distribution. For this particular data set, $\tilde{\alpha} = 0.0348$, with a corresponding 95% confidence interval (CI) $[0.0312, 0.0384]$. This means that over approximately 96% of the data are covered in the ellipsoidal set we constructed.

Note that the $\tilde{\alpha}$ above is for only one data set realization and is therefore subject to statistical fluctuation. More meaningful estimates are the averages across $1,000$ different realizations of data sets. From these we get $\hat{\alpha} = 0.0202$, and $\hat{\beta} = 0.0200$ with 95% CI $[0.0112, 0.0288]$. Notice that $\hat{\beta}$ is close to $\beta_{theoretical} = 0.0182$, which supports the validity of our algorithm.

3.1.2 Ball Uncertainty Set

When sample size is small, constructing a general ellipsoidal set might cause computation difficulty since the estimated covariance matrix could be singular. In this case, we can replace the estimated covariance

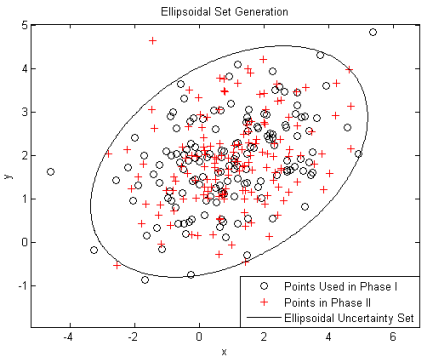

Figure 1: Ellipsoid uncertainty set.

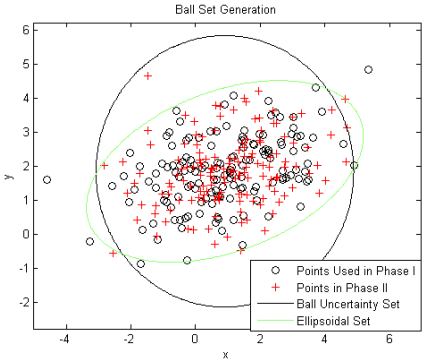

Figure 2: Ball uncertainty set.

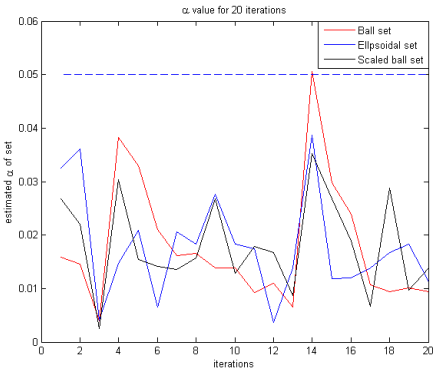

Figure 3: $\hat{\alpha}$ of three types of uncertainty set for 20 replications

matrix Σ by an identity matrix. This gives us a ball uncertainty set. In other words, the construction of a ball only needs to estimate the center of the ball, which is taken to be the sample mean.

The downside of using a ball is also obvious. Since less parameters are included, the uncertainty set could not represent the high probability region of the data as well as the general ellipsoid. This is observed in Figure 2, where the constructed ball is much larger than the general ellipsoid and covers some area where not many data points are located at.

From $1,000$ replications, we obtain $\hat{\alpha} = 0.0201$ and $\hat{\beta} = 0.0230$ with 95% CI $[0.0136, 0.0324]$. Note that these numbers are similar to those of the general ellipsoidal set. This confirms the theoretical ground discussed before that the shape of the uncertainty set does not affect its statistical validity. The similar performances between general ellipsoids and balls could also be observed from Figure 3, which shows no significant pattern of differences in the value of $\tilde{\alpha}$ across 20 different data sets.

3.1.3 Scaled Ball Uncertainty Set

Scaled ball is a compromise between general ellipsoid and ball. It reduces the data size requirement for Phase I compared to general ellipsoid, while providing a richer shape than ball. A scaled ball is an ellipsoid with center set as the sample mean, and its covariance matrix is set as a diagonal matrix where each diagonal entry is the sample variance of each dimension of the data. From a statistical perspective, using such a covariance matrix is equivalent to assuming that different dimensions of the data are independent. It can also be viewed as applying the ball set on the data standardized by the marginal variances. These two approaches lead to the same result.

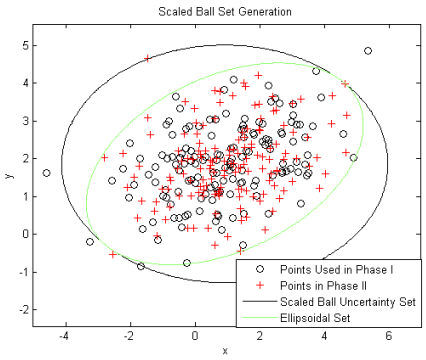

Figure 4: Scaled ball uncertainty set.

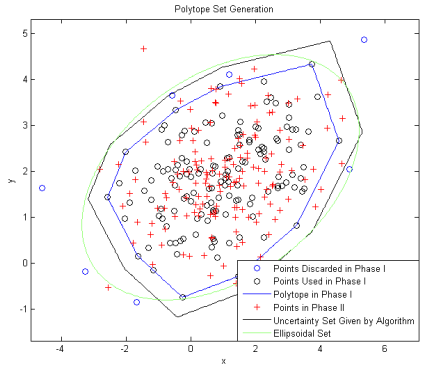

Figure 5: Polytope uncertainty set.

For the scaled ball in Figure 4, we have $\tilde{\alpha} = 0.0091$ with 95% CI $[0.0072, 0.0110]$. This value is rather small compared with the general ellipsoid and the ball we used before, indicating that the scaled ball we constructed covers more data than needed. But again, this is only for one data realization. The overall performance across $1,000$ replications, measured by $\hat{\alpha} = 0.0201$ and $\hat{\beta} = 0.0160$ with 95% CI $[0.0081, 0.0239]$, is similar to the general ellipsoid and the ball used before. The similar performance is also confirmed by Figure 3, which shows the value of $\tilde{\alpha}$ across 20 replications for all three types of sets.

Although scaled ball provides a richer model than ball, the performance of these two types of sets still depend largely on the problems to solve. For example, if the marginal variances of different dimensions are similar and the data set is not large, it is possible that ball has a better fit than scaled ball. The selection of uncertainty set geometry should be based on data exploration.

3.2 Polytope Uncertainty Set

Constructing a polytope using the convex hull of data set can be computationally intense for high-dimensional data. Comparing the tractability of RO reformulation (will be discussed more shortly), polytope is more tractable for single constraint but is less tractable than ellipsoid for multiple constraint problems.

Figure 5 shows the shape of a polytope set, whose size and coverage look similar to those of a general ellipsoidal set. The polytope in the figure has $\tilde{\alpha} = 0.0388$ with 95% CI $[0.0350, 0.0426]$. Over $1,000$ replications, we get $\hat{\alpha} = 0.0204$ and $\hat{\beta} = 0.0190$ with 95% CI $[0.0105, 0.0275]$. These statistics are similar to those of the ellipsoidal sets.

4 STRATEGY FOR SPLITTING DATA

This section presents some comparison of numerical performances among different data sizes for Phases I and II. As we have discussed previously, Phase I data are used to calibrate the shape of the uncertainty set, and its geometric type and the sample size at this phase affects the optimality level of the RO in approximating the CCP. Before we do our comparison, we first state some tractability results in implementing the RO.

4.1 Tractability of Uncertainty Sets

We first focus on a single linear chance constraint, whose RO relaxation is a robust linear program. In this case, both ellipsoidal and polytope uncertainty sets can be reformulated into tractable optimization problems.

Theorem 1 (Ben-Tal and Nemrovski 1999). Given $c \in \mathbb{R}^d$ and $b \in \mathbb{R}$, the RO

$$\begin{array}{ll} \min & c'x \\ \text{subject to} & a'x \le b \quad \forall a \in \mathscr{U} \end{array} \tag{4}$$

where $\mathscr{U} = \{a = \bar{a} + \Delta u : \|u\|_2 \leq \rho\}$ for some fixed $\bar{a} \in \mathbb{R}^d$, $\Delta \in \mathbb{R}^{d\times d}$, $\rho \in \mathbb{R}$, for $u \in \mathbb{R}^d$, is equivalent to

$$\begin{array}{ll} \min & c'x \\ \text{subject to} & a'x + \rho\|\Delta x\|_2 \leq b. \end{array}$$

Following our previous notation, $\xi = a$ is the random object, $\rho = \sqrt{s} \in \mathbb{R}$ and $\Delta = \sqrt{\Sigma} \in \mathbb{R}^{d\times d}$ for an ellipsoid $(\xi - \bar{a})'\Sigma^{-1}(\xi - \bar{a}) \leq s$. As discussed in Section 3.1, the matrix Σ can be the sample covariance matrix, diagonized covariance, or the identity matrix.

On the other hand, we have

Theorem 2 (Bertsimas et al. 2011). Given $c \in \mathbb{R}^d$ and $b \in \mathbb{R}$, the RO in (4) where $\mathscr{U} = \{a \in \mathbb{R}^d : Da \leq e\}$ for fixed $D \in \mathbb{R}^{l\times d}$, $e \in \mathbb{R}^l$ is equivalent to

$$\begin{array}{ll} \min & c'x \\ \text{subject to} & p'e \leq b \\ & p'D = x \\ & p \geq 0. \end{array}$$

where $x \in \mathbb{R}^d$ and $p \in \mathbb{R}^l$ are the decision variables.

For multiple linear joint chance constraints with dependence among the constraints, only ellipsoidal set is tractable. More precisely, for a robust linear constraint $Ax \leq b$ with uncertain $A \in \mathbb{R}^{l\times d}$, we can construct an ellipsoidal set on the vectorization of the matrix A.

Theorem 3 (Bertsimas et al. 2004). Given $c \in \mathbb{R}^d$ and $b \in \mathbb{R}^l$, the RO

$$\begin{array}{ll} \min & c'x \\ \text{subject to} & Ax \leq b \quad \forall A \in \mathscr{U} \end{array}$$

where

$$\mathscr{U} = \{A : \|M(vec(A) - vec(\bar{A}))\|_2 \leq \rho\},$$

for fixed $\bar{A} \in \mathbb{R}^{l\times d}$, $M \in \mathbb{R}^{ld}$, $\rho \in \mathbb{R}$, and $vec(A)$ as the vectorization of A, is equivalent to

$$\begin{array}{ll} \min & c'x \\ \text{subject to} & \bar{A}_i'x + \rho\|(M')^{-1}x_i\|_2 \leq b_i, \quad i = 1, ..., m, \end{array}$$

where $\bar{A}_i \in \mathbb{R}^d$ is the i-th row of $\bar{A}$, and $x_i \in \mathbb{R}^{(ld)\times 1}$ contains $x \in \mathbb{R}^d$ in entries $(i-1)d+1$ through id and 0 elsewhere.

Using our previous notation, the random object ξ is a random matrix A in this case.

The joint CCP we consider in this section is linear with 11 variables and 15 constraints:

$$\begin{array}{ll} \min & c'x \\ \text{subject to} & P(Ax \leq b) \geq 1 - \alpha \\ & x \geq 0 \end{array}$$

where $A \in \mathbb{R}^{l \times d}$ is a random matrix generated using $vec(A) \sim \mathcal{N}(vec(\bar{A}), \Sigma)$, with Σ created by multiplying a random matrix by its own transpose, and

$$\bar{A} = \begin{bmatrix} 14 & 18 & 15 & 14 & 14 & 10 & 15 & 18 & 26 & 15 & 34 \\ 98 & 24 & 14 & 14 & 14 & 0 & 0 & 30 & 0 & 0 & 0 \\ 40 & 4 & 14 & 14 & 14 & 0 & 40 & 0 & 30 & 0 & 0 \\ 28 & 6 & 13 & 14 & 15 & 55 & 0 & 20 & 0 & 20 & 0 \\ 22 & 4 & 13 & 15 & 12 & 0 & 55 & 0 & 0 & 50 & 0 \\ 22 & 4 & 13 & 13 & 12 & 0 & 0 & 55 & 0 & 0 & 0 \\ 22 & 6 & 13 & 13 & 12 & 0 & 10 & 0 & 55 & 0 & 0 \\ 26 & 58 & 15 & 15 & 12 & 0 & 0 & 10 & 0 & 55 & 0 \\ 6 & 40 & 13 & 0 & 0 & 0 & 0 & 0 & 30 & 0 & 55 \\ 6 & 12 & 13 & 13 & 26 & 0 & 0 & 0 & 0 & 0 & 0 \\ 6 & 18 & 12 & 12 & 12 & 0 & 20 & 0 & 0 & 10 & 0 \\ 6 & 18 & 15 & 14 & 14 & 0 & 0 & 0 & 10 & 0 & 0 \\ 16 & 0 & 0 & 12 & 35 & 50 & 50 & 0 & 0 & 0 & 0 \\ 20 & 0 & 10 & 36 & 50 & 0 & 0 & 50 & 50 & 0 & 0 \\ 16 & 0 & 0 & 12 & 35 & 0 & 10 & 0 & 0 & 50 & 50 \end{bmatrix} \quad b = \begin{bmatrix} 5200 \\ 4150 \\ 3550 \\ 3450 \\ 1600 \\ 950 \\ 3800 \\ 3950 \\ 2700 \\ 2300 \\ 3100 \\ 4200 \\ 5200 \\ 5200 \\ 3600 \end{bmatrix} \quad c = \begin{bmatrix} -139 \\ -88 \\ -133 \\ -137 \\ -165 \\ -80 \\ -120 \\ -150 \\ -100 \\ -150 \\ -110 \end{bmatrix}.$$

Even though this is a specific example, its numerical results will provide some general insights. As in Section 3, we perform $1,000$ replications for each setting. For each replication, we compute $\tilde{\alpha} = \hat{P}(A \notin \mathscr{U})$ from $10,000$ independently generated data points to approximate the coverage of $\mathscr{U}$. We output the sample mean and standard deviation of $\tilde{\alpha}$ across $1,000$ replications, and also $\hat{\beta} = \hat{P}(\hat{\alpha} > \alpha)$ as an estimate of the confidence level. In addition, we also look at the mean and standard deviation of the optimal values (o.v.) across these replications.

4.2 Phase I Data

We first investigate the effect of sample size in Phase I. Fixing $n_2 = 554$ and using different values for n_1, we run our experiments using ball and general ellipsoid. The results are in Table 1.

n_2=554

	n_1	o.v. mean	o.v. std
Ball	15	-2720.9	147.0181
	200	-2782.6	99.1025
	500	-2786.4	102.4721
Ellipsoid	n_1	o.v. mean	o.v. std
	200	-1108.5	69.5881
	500	-2488.9	79.0381
	4000	-2788.7	98.4885

Table 1: Optimality performances when fixing n_2 and changing n_1.

Note that using ball sets gives much better performances than general ellipsoids in the case of small sample. General ellipsoids only catch up the performance of balls for $n_1 = 15$ when n_1 is large as 4000. This can be explained by the fact that constructing a ball only requires estimating its center, whereas constructing a general ellipsoid requires additionally estimating a covariance matrix, which has a lot more parameters. With small sample, estimation of the covariance matrix tends to be poor, thus downgrading the capability of the ellipsoid to cover a high probability region of the distribution. On the other hand, however, the room for improvement is not significant for a ball as we increase n_1, as shown by the similar o.v. means for $n_1 = 200$ and 500. This is because ball is a simple geometry that is parametrized by only few parameters.

4.3 Phase II Data

We investigate the effect of sample size in Phase II. Note that for a given n_2, we can compute, using the argument in Lemma 1,

$$\beta_{theoretical} = P_D(P(A \in \mathscr{U}) \geq 1-\alpha) = 1 - \sum_{k=0}^{r^*-1} \binom{n}{k} (1-\alpha)^k \alpha^{n-k}$$

which represents the theoretical confidence level in using our procedure. Figures 6 and 7 illustrate how $\beta_{theoretical}$ changes with n_2.

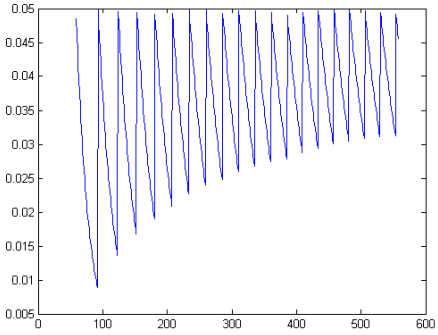

Figure 6: $\beta_{theoretical}$ against n_2 when $\beta = 0.05$ and $\alpha = 0.05$

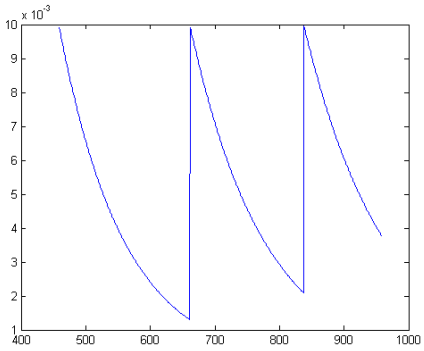

Figure 7: $\beta_{theoretical}$ against n_2 when $\beta = 0.01$ and $\alpha = 0.01$

Note that $\beta_{theoretical}$ is not strictly increasing with n_2. In the case $\beta = 0.05$ and $\alpha = 0.05$, local maximum of $\beta_{theoretical}$ occurs at n_2 =59, 93, 124 ,153, 181... In theory, these are the preferred values of n_2. Table 2 shows how n_2 affects the optimal solution, with fixed $n_1 = 500$.

n_1=500

n_2	$\hat{\beta}$	$\beta_{theoretical}$	mean of $\tilde{\alpha}$	std of $\tilde{\alpha}$	o.v. mean	o.v. std
59	0.039	0.0485	0.0159	0.0155	-2402.3	101.0091
124	0.047	0.0495	0.0241	0.0137	-2446.1	86.4363
312	0.043	0.0484	0.0319	0.0097	-2475.5	83.135
553	0.037	0.0312	0.0343	0.0079	-2481	81.2055
554	0.057	0.0491	0.0362	0.0083	-2488.9	79.0381
555	0.05	0.0482	0.0361	0.0078	-2487.4	81.2099

Table 2: Optimality and feasibility performances when fixing n_1 and changing n_2.

As n_2 increases, $\tilde{\alpha}$ has a trend to go closer to 0.05, which leads to a better solution. However, if we compare the last 3 rows, we observe that the local maximum of $\beta_{theoretical}$ (i.e., $n_2 = 554$) we mentioned above gives better results both on the mean of $\tilde{\alpha}$ and the optimal value.

4.4 Balance of Two Phases

We now fix the total sample size N and investigate the balance between n_1 and n_2. We first set $N = 500$, and use ball sets. Table 3 shows our performances using different partitions of n_1 and n_2.

Since only few parameters in the center needs to be estimated in Phase I, the importance of n_2 dominates n_1 in this case. It appears that $n_1 = 66$ and $n_2 = 434$ gives the best solution among all.

We next increase N to $5,000$, and use general ellipsoidal sets. Table 4 shows that generally more n_1 is preferable, which coincides with our intuition that more data now supports better shape estimation and hence better solutions. As a comparison, using a ball set with $n_1 = 500$ and $n_2 = 4500$ only gives an optimal solution with mean value -2.90E+03.

n_1	n_2	$\hat{\beta}$	$\beta_{theoretical}$	mean of $\tilde{\alpha}$	std of $\tilde{\alpha}$	o.v. mean	o.v. std
400	100	0.033	0.0371	0.0189	0.013	-2.49E+03	237.699
375	125	0.068	0.0478	0.0248	0.0144	-2.60E+03	220.9415
240	260	0.055	0.0499	0.0309	0.0105	-2.71E+03	143.3216
164	336	0.048	0.0487	0.0326	0.01	-2.73E+03	131.694
66	434	0.055	0.0488	0.0345	0.0091	-2.75E+03	119.9492
18	482	0.059	0.0488	0.0356	0.0088	-2.73E+03	134.5421

Table 3: $N = 500$ using ball set

n_1	n_2	$\hat{\beta}$	$\beta_{theoretical}$	mean of $\tilde{\alpha}$	std of $\tilde{\alpha}$	o.v. mean	o.v. std
4740	260	0.051	0.0499	0.03	0.0108	-3.03E+03	39.9961
4000	1000	0.054	0.0433	0.0389	0.0067	-3.04E+03	35.8851
3000	2000	0.058	0.0423	0.0418	0.0049	-3.02E+03	39.4966
2000	3000	0.066	0.0488	0.0436	0.0043	-2.98E+03	48.1733

Table 4: $N = 5,000$ using general ellipsoid set

5 COMPARISON WITH SCENARIO GENERATION APPROACH

In this section we further illustrate the strength of our approach under a small dataset. We use a linear programming problem that has only a single chance constraint and d decision variables. We use a single chance constraint because the optimal solution and the left-hand side of the constraint for any solution can be evaluated easily. This makes it convenient to compare the performance of our approach and the scenario generation (SG) approach of Calafiore and Campi (2005). In addition, we also compare the performance of different types of uncertainty sets used in our approach.

We assume that there are only $N = 120$ data points and the problem is

$$\begin{array}{ll} \min & c'x \\ \text{subject to} & P(a'x \le b) \ge 1-\alpha \end{array}$$

where $a \sim \mathcal{N}(\bar{a}, \Sigma)$, Σ is created by multiplying a random matrix by its own transpose, and c is set initially by randomly generated as $c_i = 20(1+\varepsilon_i)$, where $\varepsilon_i \sim N(0,1)$ and $a = -c$, $b = 1200$. In addition, we set $\alpha = \beta = 0.05$, and we split the data equally between the first and second phases, i.e., $n_1 = n_2 = 60$.

We consider three instances of the problem, each with a different dimension d, i.e., $d = 11, 50, 100$, and for each instance, we use independently generated data to solve the problem 1,000 times. We use the average and standard deviation of the optimal value as the measure of the "optimal level" of each algorithm. We then compute the left-hand side of the chance constraint at the obtained solution x_u, i.e., $\tilde{\alpha} = P(a'x_u > b)$ (note that this definition of $\tilde{\alpha}$ is different from in Sections 3 and 4, and captures the conservativeness relative to the solution), and estimate the constraint violation probability, i.e., $\tilde{\beta} = P_D(\tilde{\alpha} > \alpha)$. Note that, by the definitions of α and β, we expect $\tilde{\alpha} \le \alpha$ and $\tilde{\beta} \le \beta$ for our approach. We also compute $\tilde{\alpha}$ and $\tilde{\beta}$ of the SG approach with the given dataset to see whether they satisfy $\tilde{\alpha} \le \alpha$ and $\tilde{\beta} \le \beta$. For each instance, we run our approach 1,000 times, each with an independently generated dataset, and report the estimated mean and standard deviation of the optimal values (o.v.), the estimated mean and standard deviations of $\tilde{\alpha}$, and the estimated $\tilde{\beta}$ and its 95% lower and upper confidence bounds in corresponding tables (Tables 5, 6 and 7 correspond to the instances $d = 11$, 50 and 100).

We first consider the instance where $d = 11$, the true optimal value of the problem is -1196.7, and the results are reported in Table 5. Because the sample size is not sufficient for the SG approach to deliver the desired α and β levels, we observe that its average $\tilde{\alpha}$ is above $\alpha = 0.05$ and its solution violates the chance constraint with a high probability $\tilde{\beta}$. Among the three types of uncertainty sets used in our approach, the ellipsoidal set has the best performance, while the scaled ball and ball sets are more conservative. As well known in the literature, RO approach gives conservative solutions. This also appears in our results, as both $\tilde{\alpha}$ and $\tilde{\beta}$ are close to zero, indicating a high level of conservativeness.

	SG	ellipsoid	scaled ball	ball
o.v. mean	-1196.60	-1189.30	-946.25	-934.12
o.v. std	0.465	1.456	31.918	31.031
$\tilde{\alpha}$ mean	0.0924	1.27E-05	1.85E-04	1.68E-04
$\tilde{\alpha}$ std	0.0277	4.58E-05	5.90E-04	5.49E-04
$\tilde{\beta}$	0.9620	0	0	0
$\tilde{\beta}$ CI lower	0.9500	0	0	0
$\tilde{\beta}$ CI upper	0.9740	0	0	0

Table 5: The instance with $d = 11$.

	SG	ellipsoid	scaled ball	ball
o.v. mean	-1197.90	-1152.05	-924.31	-922.93
o.v. std	0.695	12.847	13.485	12.510
$\tilde{\alpha}$ mean	0.4108	1.18E-04	5.07E-17	3.87E-17
$\tilde{\alpha}$ std	0.0402	5.46E-04	7.19E-16	7.02E-16
$\tilde{\beta}$	1.0000	0	0	0
$\tilde{\beta}$ CI lower	1.0000	0	0	0
$\tilde{\beta}$ CI upper	1.0000	0	0	0

Table 6: The instance with $d = 50$.

	SG	scaled ball	ball
o.v. mean	-	-831.50	-828.15
o.v. std	-	11.582	11.253
$\tilde{\alpha}$ mean	-	0	0
$\tilde{\alpha}$ std	-	0	0
$\tilde{\beta}$	-	0	0
$\tilde{\beta}$ CI lower	-	0	0
$\tilde{\beta}$ CI upper	-	0	0

Table 7: The instance with $d = 100$.

We now increase d to 50, and the true optimal value is -1195.01. Among the 1,000 replications of the experiments, SG approach detects that its sample problem is unbounded for 28 times. Even though we eliminate those unbounded cases in the results reported in Table 6, the solution given by SG still violates the chance constraint in every replications, i.e., $\tilde{\beta} = 1$. In the meanwhile, the solutions found by our approach satisfy the constraint and the performance is similar to the instance of $d = 11$.

When $d = 100$, the true optimal value is -1196.32. For this instance, SG approach encounters unbounded sample problems in all $1,000$ replications. Therefore, we have no results to report in Table 7 for SG approach. Because the dimension is too large for 60 Phase I data to give a valid estimate of the covariance matrix, we only apply the scaled ball and ball sets in our approach. Again, our approach finds solutions that satisfy the constraint but are conservative.

ACKNOWLEDGMENTS

The research of L. Jeff Hong was partially supported by the Hong Kong Research Grants Council under projects GRF613213 and GRF16203214. The research of Zhiyuan Huang and Henry Lam was partially supported by the National Science Foundation under grants CMMI-1400391/1542020 and CMMI-1436247/1523453.

REFERENCES

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

Ben-Tal, A., and A. Nemkrovski. 1999. "Robust solutions of uncertain linear programs". *Operations Research Letters* 25 (1):1-13.

Bertsimas, D., D. B. Brown, and C. Caramanis. 2011. "Theory and applications of robust optimization". *SIAM Review* 53 (3): 464-501.

Bertsimas, D., D. Pachamanova, and M. Sim. 2004. "Robust linear optimization under general norms". *Operations Research Letters* 32 (6): 510-516.

Boyd, S. and L. Vandenberghe. 2004. *Convex Optimization*, Cambridge University Press, Cambridge, UK.

Calafiore, G. and M. C. Campi. 2005. "Uncertain convex programs: Randomized solutions and confidence levels". *Mathematical Programming* 102: 25-46.

Calafiore, G. and M. C. Campi. 2006. "The scenario approach to robust control design". *IEEE Transactions on Automatic Control* 51: 742-753.

Charnes, A., W. W. Cooper, and G. H. Symonds. 1958. "Cost horizons and certainty equivalents: An approach to stochastic programming of heating oil". *Management Science* 4: 235-263.

Dentcheva, D., B. Lai, and A. Ruszczyński. 2004. "Dual methods for probabilistic optimization problems". *Mathematical Methods of Operations Research* 60: 331-346.

Hallin, M., D. Paindaveine, M. Šiman, Y. Wei, R. Serfling, Y. Zuo, L. Kong, and I. Mizera. 2010. "Multivariate quantiles and multiple-output regression quantiles: From L_1 optimization to halfspace depth [with Discussion and Rejoinder]". *Annals of Statistics* 38 (2): 635-703.

Hong, L. J., Y. Yang, and L. Zhang. 2011. "Sequential convex approximations to joint chance constrained programs: A Monte Carlo approach". *Operations Research* 59: 617630

Luedtke, J., S. Ahmed, and G. Nemhauser. 2007. "An integer programming approach for linear programs with probabilistic constraints". *Lecture Notes on Computer Science* 4513: 410-423.

Miller, L. B. and H. Wagner. 1965." Chance-constrained programming with joint constraints". *Operations Research* 13:930-945.

Prékopa, A. 1970. "On probabilistic constrained programming". *Proceedings of the Princeton Symposium on Mathematical Programming*, 113-138.

Prékopa, A. 2003. "Probabilistic programming". In *Stochastic Programming, Handbooks in OR&MS*. Vol. 10, A. Ruszczynski and A. Shapiro, eds., Elsevier.

Prékopa, A., T. Rapcsák, and I. Zsuffa. 1978. "Serially linkedreservoir system design using stochastic programming". *Water Resources Research* 14: 672-678.

Serfling, R. 2002. "Quantile functions for multivariate analysis: approaches and applications". *Statistica Neerlandica* 56 (2): 214-232.

Shi, Y., J. Zhang, and K. B. Letaief. 2015. "Optimal stochastic coordinated beamforming for wireless cooperative networks with CSI uncertainty". *IEEE Transactions on Signal Processing* 63(4): 960-973.

AUTHOR BIOGRAPHIES

L. JEFF HONG received his Ph.D. in Industrial Engineering and Management Science from Northwestern University, and is currently a Chair Professor of Management Sciences at City University of Hong Kong. His research interests include stochastic optimization and stochastic simulation, financial engineering and risk management, and business analytics. His email address is jeffhong@cityu.edu.hk.

ZHIYUAN HUANG is a first year Ph.D. student in the Department of Industrial and Operations Engineering at the University of Michigan, Ann Arbor. His research interests include simulation and stochastic optimization. His email address is zhyhuang@umich.edu.

HENRY LAM is an Assistant Professor in the Department of Industrial and Operations Engineering at the University of Michigan, Ann Arbor. His research focuses on stochastic simulation, risk analysis, and simulation optimization. His email address is khlam@umich.edu.